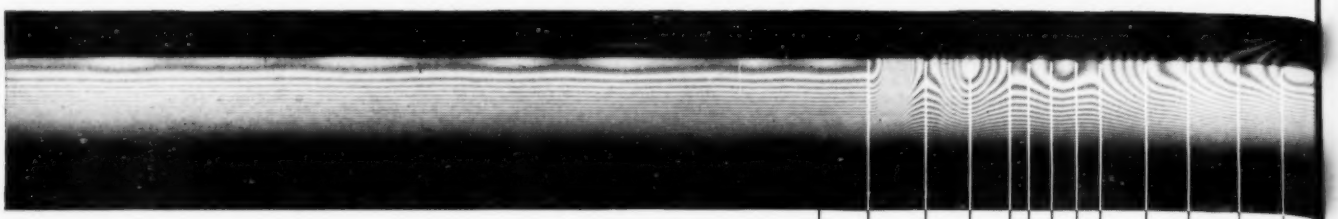
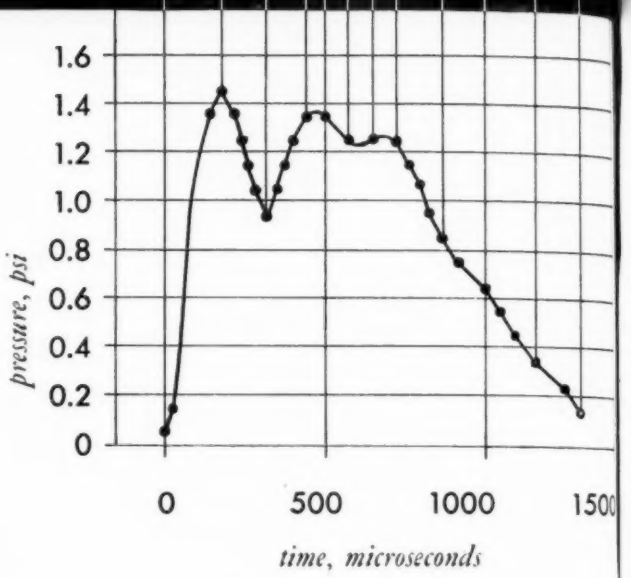
JOURNAL OF APPLIED PHYSICS

FAST WORK BY

NEWTON RINGS

capturing pressure transients





Light, the ideal yardstick for measuring rapid events, creates a pressure-transient measuring system of unusual capability in the Beckman & Whitley Model 216 Pressure Recorder. Having stable linear calibration, the unit can cover pressure ranges from 3 to 50,000 psi, with a 1 per cent full-scale accuracy in both pressure and time, and exhibits a frequency response from 10,000 to 50,000 cps.

To cover the widest possible variety of application requirements, the recorder is engineered so it can be used in three ways: (1) exposed to free air manifestations, (2) attached directly to a pressure vessel, or (3) fed by a hydraulic line from the pressure source.

The instrument is organized into two units, shown below. Transducer, optical system, and camera with timing-pip system are in the unit at left. The control unit, right, permits records to be made remotely.

Rapid response, accurate measurement of small displacements, high sensitivity, and stable zero reference points are all achieved by recording on moving film the motion of small arcs of optical-interference rings responding to the influence of the pressure fluctuations under measurement. Such fringe changes can be seen in the typical record above, while underneath, the simple procedure of translating the pattern into a plot of pressure-versus-time is outlined.

SEND FOR FURTHER DETAILS.

Beckman & Whitley INC.

instruments for physical research

SAN CARLOS 8, CALIFORNIA

See these units at Booth C-515-1, 10th Annual Instrument-Automation Conference and Exhibit, Los Angeles, September 12 to

JOURNAL OF APPLIED PHYSICS

Vol. 26, No. 6 JUNE, 1955

In This Issue

Page

Wide-Angle Radiation in Pencil Beam Antennas . . . . . David Carter 645
Effect of Reactor Irradiation on the White-to-Grey Tin Transformation . . Jerome Fleeman and G. J. Dienes 652
Monte Carlo Calculation of Gamma-Ray Albedos of Concrete and Aluminum . . . . . J. F. Perkins 655
Reverse Current and Carrier Lifetime as a Function of Temperature in Germanium Junction Diodes . . . . . E. M. Pell 659
Theoretical and Experimental Study of Electromagnetic Scattering by Two Identical Conducting Cylinders . . . . . R. V. Row 666
Trajectory-Tracer Study of Helix- and Band-Type Postdeflection Acceleration . . . . George C. Sponsler 676
Performance of the Nonmagnetic Radio-Frequency Mass Spectrometer Tube . . . . . T. C. Wherry and F. W. Karasek 682
Silicon *n-p-n* Grown Junction Transistors . M. Tanenbaum, L. B. Valdes, E. Buehler, and N. B. Hannay 686
Magnetic Materials for Digital Computer Components. II. Magnetic Characteristics of Ultra-Thin Molybdenum-Permalloy Cores . . . . . Norman Menyuk 692
Transition Through a Contact Region . . . . . J. G. Hall 698
Short-Time Stress Relaxation Behavior of Plastics . M. T. Watson, W. D. Kennedy, and G. M. Armstrong 701
Microwave Diffraction by Apertures of Various Shapes . . . . . S. J. Buchsbaum, A. R. Milne, D. C. Hogg, G. Bekefi, and G. A. Woonton 706
Noise Measurements in Semiconductors at Very Low Frequencies . . . . . T. E. Firle and H. Winston 716
Calculation of the Space Charge, Electric Field, and Free Carrier Concentration at the Surface of a Semiconductor . . . . . Robert H. Kingston and Siegfried F. Neustadter 718
Electrical Breakdown in High Vacuum . . . . . W. S. Boyle, P. Kisliuk, and L. H. Germer 720
X-Ray Measurements of Particle Size and Strain Distribution in Cold Worked Silver . . . . . F. R. L. Schoening and J. N. van Niekerk 726
Energy Relations in Cold Working an Alloy at 78°K and at Room Temperature . . . . . J. S. L. Leach, E. G. Loewen, and M. B. Bever 728
Work Function of Tungsten Single Crystal Planes Measured by the Field Emission Microscope . . . . . Erwin W. Müller 732
Viscoelasticity of Poly Methacrylates . . . . . F. Bueche 738
Experimental Examination of the Statistical Theory of Rubber Elasticity. Low Extension Studies . . . . . F. P. Baldwin, J. E. Ivory, and R. L. Anthony 750
Thermionic Emission Microscopy of Metals. 1. General . . . . . R. D. Heidenreich 757
Numerical Solutions of Spherical Blast Waves . . . . . Harold L. Brode 766

Letters to the Editor:

Method for Producing High-Velocity Metallic and Plastic Pellets . . . . . M. E. Van Valkenburg and Charles D. Hendricks, Jr. 776
Collapsed Toroidal Springs as Contact Elements in Microwave Cavities . . . C. L. Andrews 777
Further Remarks on the Odd Ball Problem as an Example in Information Theory . . . . . F. H. Mitchell and Robert N. Whitehurst 778
Potential Distribution as a Function of Current in the Spherical Diode . . . . . Annija Dunkuls and Earl Zwicker 779
Measurement of Wall Temperature and Heat Flow in the Shock Tube . . . . . Albert J. Chabai and Raymond J. Emrich 779
Generalizations of Brillouin Flow . . . . . L. R. Walker 780
Time Dispersion of Secondary Electron Emission . . . . . E. W. Ernst and H. VonFoerster 781
Ionic Pumping Mechanism of Helium in an Ionization Gauge . . . . . L. J. Varnerin, Jr., and J. H. Carmichael 782
Erratum: Analog Methods for Study of Transient Flow in Solids with Temperature-Dependent Thermal Properties . . . . . Norman E. Friedmann 783

Announcements 715
Books Reviewed 783
Books Received 784
Research Reports 784
Index to Advertisers xx

JOURNAL OF APPLIED PHYSICS

A journal of applications of physics to other sciences and to industry

Published by the

American Institute of Physics

OFFICERS

FREDERICK SEITZ
Chairman

GEORGE B. PEGRAM
Treasurer

GOVERNING BOARD

ALLEN V. ASTIN
ROBERT F. BACHER
H. A. BETHE
J. W. BUCHTA
S. A. GOUDSMIT
DEANE B. JUDD
HUGH S. KNOWLES
W. H. MARKWOOD, JR.
WILLIAM F. MEGGERS
PHILIP M. MORSE
BRIAN O'BRIEN
HARRY F. OLSON
R. F. PATON
GEORGE B. PEGRAM
ERIC RODGERS
RALPH A. SAWYER
FREDERICK SEITZ
WILLIAM SHOCKLEY
H. D. SMYTH
J. H. VAN VLECK
MARK W. ZEMANSKY

Consultant on Printing

MELVIN LOOS

ADMINISTRATIVE STAFF

HENRY A. BARTON
Director

WALLACE WATERFALL
Executive Secretary

THEODORE VORBURGER
Advertising Manager

RUTH F. BRYANS
Publication Manager

EDITH I. NEFTEL
Circulation Manager

KATHRYN SETZE
Assistant Treasurer

Editor
Robert L. Sproull

Editorial Board

Maurice A. Biot
Walker Bleakney
Wallace J. Eckert
Elmer Hutchisson
D. O. North
Louis A. Pipes
W. R. Smythe
Arthur V. Tobolsky
A. M. Weinberg

THE JOURNAL OF APPLIED PHYSICS is devoted to the application of physics to other sciences, to engineering, and to industry. Previous to 1937, this journal was called PHYSICS.

The JOURNAL OF APPLIED PHYSICS welcomes articles of two distinct kinds: (1) Descriptions of new results in experimental or theoretical physics which are of general interest to applied physicists. The author should explain his contribution in such a way that a physicist or engineer in another field of specialization can appreciate its objectives and essential features. (2) Critical reviews of particular areas of applied physics. Authors contemplating writing such reviews should correspond with the Editor before they begin writing.

The JOURNAL OF APPLIED PHYSICS welcomes letters to the editor which describe in a few words the content of laboratory reports and theses and which tell readers how they may obtain copies of such reports.

Manuscripts should be submitted to the Editor, Rockefeller Hall, Cornell University, Ithaca, New York. Authors should read the "Information for Contributors," facing page 1 of the January issue, before writing articles or letters. The *Style Manual* of the American Institute of Physics is also helpful to writers preparing manuscripts for journals published by the American Institute of Physics. It can be obtained for $1.00 from the American Institute of Physics, 57 East 55th Street, New York 22, New York.

Corrected proofs and all correspondence concerning papers in the process of publication should be addressed to the Publication Manager, American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	14.00

Back Number Prices

Yearly back number rate when complete year is available: $15.00.
Single copies: $1.50 each.

Changes of address should be sent to the American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 26, Number 6 June, 1955

Wide-Angle Radiation in Pencil Beam Antennas

DAVID CARTER
Convair Division, General Dynamics Corporation, San Diego, California
(Received July 16, 1954)

This paper is concerned with theoretical estimates of wide-angle radiation in pencil beam antennas. In reflector-type antennas this energy consists of direct radiation from the feed and scattered energy from the reflector. Approximate methods for the evaluation of these contributions are discussed, together with their simplifying assumptions. To get some numerical indications, calculations were made for paraboloidal reflectors of different f/D ratios and a class of primary patterns which provide an approximate representation of a great many common feeds. The results are presented in graphical form to provide useful design information. They show the dependence of the wide-angle radiation level on feed and dish parameters.

IN many applications of pencil beam antennas it is important to know the amount of energy directed at large angles away from the axis of the main beam. While this energy is usually very small, some operations may require all wide-angle radiation to be down 50 db or more. Physical proximity of antennas may impose an extremely small upper limit on the absolute power in certain directions in order to reduce cross talk or prevent crystal burn-out. This may in turn, for certain geometries, add a very stringent specification of side lobe discrimination at large angles to the usual gain and small-angle side lobe specifications. Then feed and reflector design will require a theoretical estimate of the antenna pattern at large angles off axis.

In reflector-type antennas, the energy radiated at large angles off the beam axis consists of direct radiation from the feed and scattered energy from the reflector. In any direction where these two contributions are of the same order of magnitude a knowledge of their relative phases is necessary to obtain the power from the resultant of the two field intensities. In the usual case, however, the contribution from one far exceeds that of the other, and the smaller one can then be neglected.

It is a simple matter to approximate the contribution of the direct radiation from the feed to the secondary pattern. The power per unit solid angle in the secondary field $P_S(\theta,\phi)$ is related to the corresponding quantity in the primary field $P_P(\xi,\psi)$ by the definition of primary feed gain function $G_P(\xi,\psi)$ and secondary pattern gain function $G_S(\theta,\phi)$. Thus, for a system having rotational symmetry, as shown in Fig. 1, when the diffracted energy at θ_1 is negligible, $P_S(\theta_1)\simeq P_P(\psi_1)$. Then, since the total radiated power is the same for primary and secondary patterns,

$$\frac{P_S(\theta_1)}{P_S(0)}\simeq\frac{G_P(0)}{G_S(0)}\cdot\frac{P_P(\psi_1)}{P_P(0)}.$$

Writing this in a slightly more practical form—if N_S is the maximum allowable db level of a side lobe in the secondary pattern, produced solely by direct radiation from the primary feed as shown in Fig. 1, then the db level of the corresponding radiation in the primary feed pattern, $N_P\equiv 10\log_{10}[P_P(\psi_1)/P_P(0)]$, must be given by

$$N_P\leqq N_S+10\log_{10}\frac{G_S(0)}{G_P(0)}.$$

This indicates, what was intuitively apparent, that the larger the secondary gain and the smaller the primary feed gain, the less stringent will be the specification against primary feed spill-over energy (i.e., the feed radiation which is not intercepted by the reflector).

To get some numerical indication, this relationship was evaluated for paraboloidal reflectors of different f/D ratios and a class of primary patterns which provide an approximate representation of a great

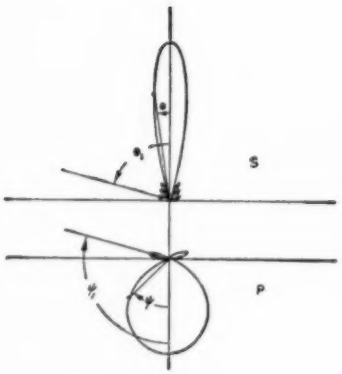


FIG. 1. Corresponding primary and secondary radiation pattern representations. The primary energy at ψ_1 is not intercepted by the reflector and it appears at θ_1 in the secondary pattern.

many of the common feeds. Assuming

$$G_P(\psi)=\begin{cases} G_P(0)\cos^n\psi & 0\leqq\psi\leqq\dfrac{\pi}{2} \\ \text{for} & \\ \epsilon\delta(\psi-\psi_1) & \dfrac{\pi}{2}\leqq\psi\leqq\pi \end{cases},$$

where $\epsilon\ll 2$, and δ is the Dirac delta function, it can be shown[1] that

$$G_S(0)=\left(\frac{\pi D}{\lambda}\right)^2 G_P(0)\left[\cot\frac{\Psi}{2}\int_0^{\Psi}\cos^{n/2}\psi\tan\frac{\psi}{2}d\psi\right]^2,$$

where D is the aperture diameter, λ the operating wavelength, and Ψ is the angular aperture. For such antennas then

$$N_P-N_S-20\log_{10}\frac{\pi D}{\lambda} \leqq 10\log_{10}\left[\cot\frac{\Psi}{2}\int_0^{\Psi}\cos^{n/2}\psi\tan\frac{\psi}{2}d\psi\right]^2.$$

This relationship has been plotted in Fig. 2 for the different feed patterns characterized by the values of the parameter $n=2$, 4, 6, and 8. It affects the choice of a reflector for a given primary feed and side lobe spec when the side lobe in the secondary pattern is due to direct radiation from the feed alone, or specifies the maximum allowable level of this radiation for a fixed reflector. For example, consider the case of a paraboloid, fixed by other considerations, with a diameter of 30 in. and an angular aperture of 70° illuminated by 3.2 cm radiation. If the feed pattern within this angular aperture is approximately $\cos^6\psi$, the $n=6$ curve in Fig. 2 fixes the maximum value of $N_P-N_S-20\log_{10}(\pi D/\lambda)$ at -14.1 db for $\Psi=70°$, and therefore, in this example, $N_P\leqq N_S+37.5-14.1=N_S+23.4$ db. Thus, if wide-angle radiation produced solely by directly radiated feed energy were required to be below $N_S=-40$ db in this case, this would require the primary pattern to be below -16.6 db for all angles ψ greater than 70°.

The other contributing source of radiation in the secondary pattern is the energy scattered by the reflector. Usually this is calculated as a diffraction pattern from a scalar distribution over the aperture.[2] Unfortunately this aperture distribution method involves approximations which, among other things, limits the applicability of the results to angles near the axis of the main lobe.

Another approach which removes this restriction to small angles and reduces to the aperture distribution formula for small angles off the main beam axis is the method of calculating the scattered pattern from the current distribution over the reflector. This formulation differs from the former only by replacing the scalar integration over the aperture by a vector integration of the equivalent surface current density over the entire surface of the reflector. However, the calculation of the far field leads to significant differences at large angles off the main lobe axis.

Using the latter approach then to obtain an estimate of the scattered field, it can be shown that the far field is given by

$$E_{\theta,\phi}\approx\text{const}\frac{e^{-jkR}}{R}\mathbf{i}_{\theta,\phi}\cdot\mathbf{I}$$

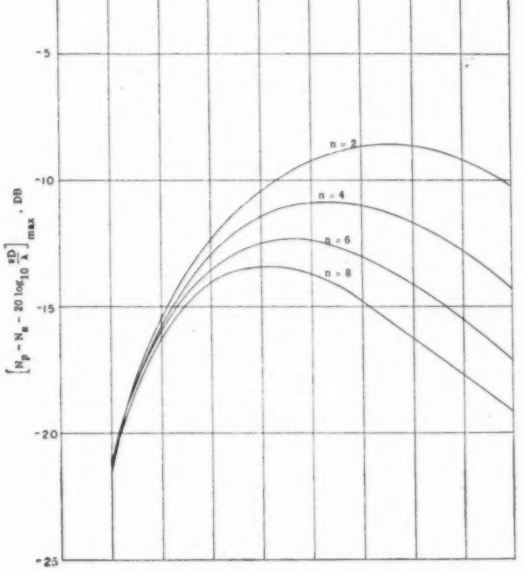


FIG. 2. Dependence on angular aperture and primary feed gain of a function relating primary and secondary pattern levels of spill-over energy in paraboloid antennas. The primary feed gain function is $G_P=2(n+1)\cos^n\psi$ in the region subtended by the paraboloid.

[1] S. Silver, *Microwave Antenna Theory and Design* (McGraw-Hill Book Company, Inc., New York, 1949), Chap. 12.

[2] S. Silver, reference 1, p. 167.

and

$$\mathbf{H}_P \approx \left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}} \mathbf{R}_1 x \mathbf{E}_P,$$

where

$$\mathbf{I} = \int_S \mathbf{N} x \mathbf{H} e^{jk\boldsymbol{\rho}\cdot\mathbf{R}_1} dS.$$

Here $\mathbf{i}_\theta$ and $\mathbf{i}_\phi$ are unit vectors at the field point P in the polar and azimuthal directions, respectively, $\boldsymbol{\rho}$ is the vector from the origin to the surface element dS, $\mathbf{R}_1$ is the unit vector pointing from the origin to the point P at which the field is being calculated, $k=2\pi/\lambda$, $j=\sqrt{-1}$, and $\mathbf{N}$ is the unit normal vector at dS; R is the distance from the origin to the field point, ϵ and μ are the electric and magnetic inductive capacities, respectively, and S denotes the reflector surface. This may be obtained from an integration of Maxwell's equations, expressing the field in terms of the sources.[3] The lowest-order terms in R for the case of a source free region bounded by one infinitely conducting surface then leads to the far-field approximation already given.

Several simplifications and assumptions are made in order to obtain workable formulas. To anticipate discrepancies between theory and experiment, it seems pertinent to review some of the assumptions and approximations that are made in the application of the current distribution formulas.

The scattered field in the far zone may be calculated once the $\mathbf{H}$ distribution over the entire surface of the reflector is known. To obtain an approximation in terms of the primary feed pattern, it is assumed that the primary feed radiation is geometrically reflected and that multiple scattering between the feed and reflector may be neglected because of their geometry and large separation. These assumptions imply that the reflector is in the far zone of the feed, and that the curvatures of both the reflector and primary feed wave front are so small that the transformation from incident to reflected wave front may be calculated, at each point of the reflector, as though the tangent plane at that point reflects a uniform plane wave having the same intensity and polarization as the incident wave front at that point. The tangential component of the total $\mathbf{H}$ at any point on the reflector is then given in terms of the reflected primary field at the same point by

$$\mathbf{N} x \mathbf{H} = 2\mathbf{N} x \mathbf{H}_r = 2\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}} \mathbf{N} x (\mathbf{S}_1 x \mathbf{E}_r),$$

where $\mathbf{S}_1$ is a unit vector in the direction of Poyntings' vector on the reflected wave front and the subscript r refers to reflected quantities. The far-field relations for a point source feed gives $\mathbf{E}_r$ (which, from the geometrical optics approximation, differs from the incident $\mathbf{E}$ at the reflector only in polarization) in terms of the primary gain function as

$$\mathbf{E}_r = \text{const} \frac{e^{-jk\rho}}{\rho} G_P^{\frac{1}{2}}(\psi,\xi)\mathbf{e}_r(\psi,\xi),$$

where $\mathbf{e}_r$ is a unit vector defining the polarization in the reflected wave front.

Substitution of these approximations then allows the vector integral $\mathbf{I}$, which determines the scattered field, to be expressed as an integral of the primary feed pattern over the surface of the reflector. Thus

$$\mathbf{I} = \text{const} \int_S \frac{G_P^{\frac{1}{2}}(\psi,\xi)}{\rho} \exp\{-jk\rho[1-\boldsymbol{\rho}_1\cdot\mathbf{R}_1]\} \times[(\mathbf{N}\cdot\mathbf{e}_r)\mathbf{S}_1-(\mathbf{N}\cdot\mathbf{S}_1)\mathbf{e}_r]dS,$$

where $\boldsymbol{\rho}_1$ is the unit vector pointing from the origin to the element dS.

The field on the shadow side of the reflector is assumed to vanish, which is a consequence of using the approximations of the geometrical optics method of calculating the field over the reflector. Normally there will be small currents at the edge and on the shadow side of the reflector which will contribute to the diffraction field, especially in the shadow region. In some cases this might be the worst assumption. However, if the edge of the reflector is very near a deep null of the primary feed pattern, these currents will be negligibly small, and this will be assumed in what follows.

To obtain numerical estimates of this second contribution to wide-angle radiation, the scattered pattern was calculated under the same conditions as in the previous calculations of the direct radiation from the feed, *viz.* paraboloidal reflectors of different f/D ratios and the previously defined class of primary patterns, which provide an approximate representation of a great many common feeds. (Generally G_P is a function of both ψ and ξ. However, for primary feed patterns having approximately the same half-power beam widths in the principal E- and H-planes, it is usually assumed that the feed pattern can be represented by a pattern, having rotational symmetry about the paraboloid axis, which is the average of the patterns in the two principal planes.)

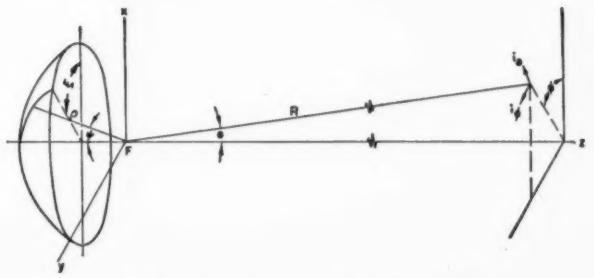


FIG. 3. Coordinate system. The origin is at the focal point of the reflector.

[3] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 467. S. Silver, reference 1, Chap. 3.

To calculate the diffraction pattern, the integral **I** is expanded in terms of the coordinates shown in Fig. 3. Thus, after two simplifying assumptions which sacrifices some range of application as discussed in the Appendix, it may be shown that for paraboloidal reflectors $[\rho = f\sec^2(\psi/2)]$,

$$\mathbf{I} = \text{const} \int_0^{\Psi} \left[\mathbf{i}_z J_0\left(2kf \sin\theta \tan\frac{\psi}{2}\right) + \mathbf{i}_x j \cos\phi \tan\frac{\psi}{2} J_1\left(2kf \sin\theta \tan\frac{\psi}{2}\right)\right] G_P^{\frac{1}{2}}(\psi) \times \exp\left[-jkf \sec^2\frac{\psi}{2}(1+\cos\theta\cos\psi)\right] \tan\frac{\psi}{2} d\psi.$$

It can be shown that in the region near the axis ($\theta \approx 0$), **I** reduces approximately to the familiar integral of the aperture distribution method of calculating the diffraction pattern. However, for wide angles, the contribution to the axial component of **I**, which has no counterpart in the aperture distribution formulation, cannot be neglected, and the two formulations give significantly different answers.

For feeds having approximately rotationally symmetric patterns, then, evaluation of this last form of **I** leads to the scattered field of the paraboloid in any direction (θ,ϕ). Restricting attention to the principal E- and H-plane patterns and changing integration variable to $x = \tan\psi/2$, it follows that, in the principal E-plane $(\phi = 0)$,

$$\mathbf{I} = \text{const} \exp\left(-j2kf\cos^2\frac{\theta}{2}\right) \int_0^{\mathrm{X}} G_P^{\frac{1}{2}}(x) \frac{x}{1+x^2} \exp\left(-j2kfx^2 \sin^2\frac{\theta}{2}\right) \times [\mathbf{i}_x J_0(2kfx\sin\theta) + \mathbf{i}_z jx J_1(2kfx\sin\theta)]dx$$

and in the principal H-plane $(\phi = \pm\pi/2)$,

$$\mathbf{I} = \text{const}\,\mathbf{i}_x \exp\left(-j2kf\cos^2\frac{\theta}{2}\right) \int_0^{\mathrm{X}} G_P^{\frac{1}{2}}(x) \frac{x}{1+x^2} \times \exp\left(-j2kfx^2\sin^2\frac{\theta}{2}\right) J_0(2kfx\sin\theta)dx.$$

Since $\mathbf{i}_\phi \cdot \mathbf{i}_z = 0$ and $\mathbf{i}_\phi \cdot \mathbf{i}_x = -\sin\phi$, it can be seen that the field in the E-plane has θ polarization, and, since $\mathbf{i}_\theta \cdot \mathbf{i}_x = \cos\theta\cos\phi$, the field in the H-plane has ϕ polarization.

For the actual calculation, the previously defined feed patterns were used, *viz.*,

$$G_P(\psi) = G(0)\cos^n\psi = G(0)\left(\frac{1-x^2}{1+x^2}\right)^n,$$

and the working formulas for the scattered field amplitude became

$$\mathbf{E}(\theta) = A \exp\left(-j2kf\cos^2\frac{\theta}{2}\right) \times \begin{cases} \mathbf{i}_\theta[I_{x2}\sin\theta - I_{x1}\cos\theta + j(I_{x1}\sin\theta + I_{x2}\cos\theta)] & \text{in the } E\text{-plane} \\ \pm\mathbf{i}_\phi(I_{x1} - jI_{x2}) & \text{in the } H\text{-plane} \end{cases}$$

where

$$I_{x1} = \int_0^{\mathrm{X}} \left(\frac{1-x^2}{1+x^2}\right)^{n/2} \frac{x}{1+x^2} \cos\left(2kfx^2\sin^2\frac{\theta}{2}\right) J_0(2kfx\sin\theta)dx,$$

$$I_{x2} = \int_0^{\mathrm{X}} \left(\frac{1-x^2}{1+x^2}\right)^{n/2} \frac{x}{1+x^2} \sin\left(2kfx^2\sin^2\frac{\theta}{2}\right) \times J_0(2kfx\sin\theta)dx,$$

$$I_{z1} = \int_0^{\mathrm{X}} \left(\frac{1-x^2}{1+x^2}\right)^{n/2} \frac{x^2}{1+x^2} \cos\left(2kfx^2\sin^2\frac{\theta}{2}\right) \times J_1(2kfx\sin\theta)dx,$$

$$I_{z2} = \int_0^{\mathrm{X}} \left(\frac{1-x^2}{1+x^2}\right)^{n/2} \frac{x^2}{1+x^2} \sin\left(2kfx^2\sin^2\frac{\theta}{2}\right) \times J_1(2kfx\sin\theta)dx,$$

and A is a function of R alone. To normalize the pattern, the maximum value of E,

$$E_{max} = |AI_{x1}(0)| = \left|A\int_0^{\mathrm{X}} \left(\frac{1-x^2}{1+x^2}\right)^{n/2} \frac{x}{1+x^2}dx\right|,$$

is easily evaluated for any value of n. Substituting in the power pattern equation

$$r = 10\log_{10}\left(\frac{E}{E_{max}}\right)^2$$

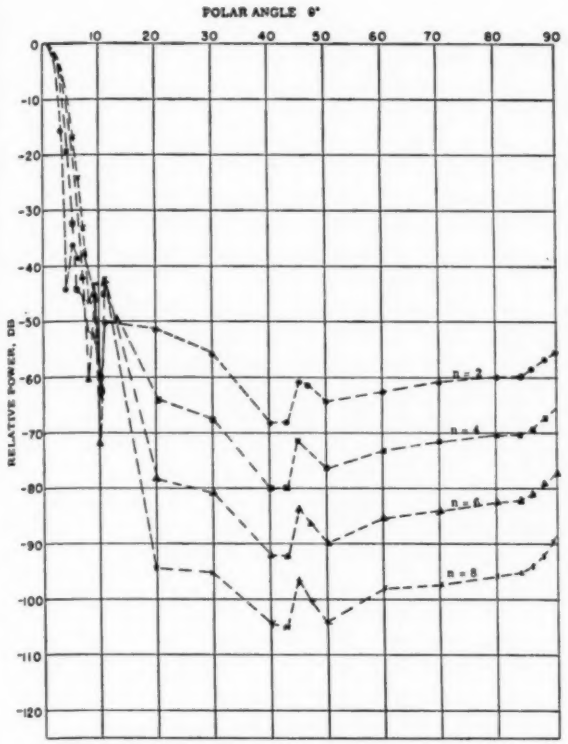
then gives the diffraction pattern. Thus, in the E-plane,

$$r_E = 10\log_{10}\frac{1}{I_{x1}^2(0)}[(I_{x1}^2 + I_{x2}^2)\cos^2\theta + (I_{x1}^2 + I_{x2}^2)\sin^2\theta + (I_{x1}I_{x2} - I_{x1}I_{x2})\sin 2\theta]$$

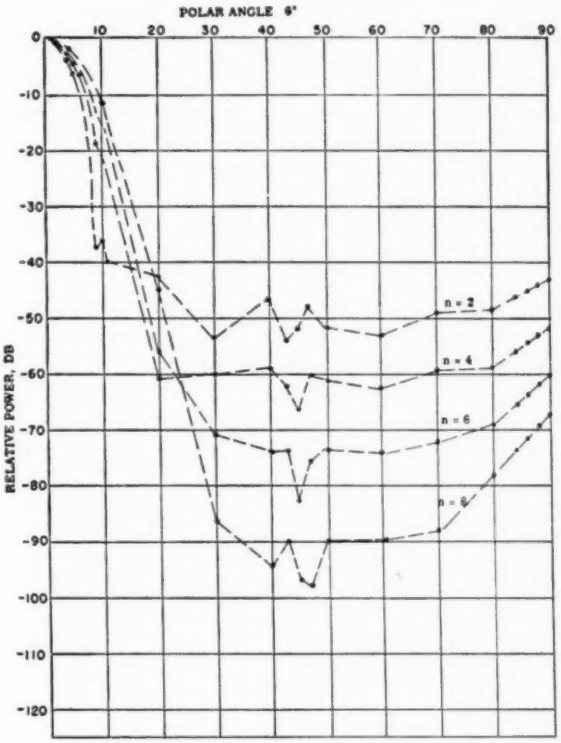
and in the H-plane,

$$r_H = 10\log_{10}\frac{1}{I_{x1}^2(0)}[I_{x1}^2 + I_{x2}^2].$$

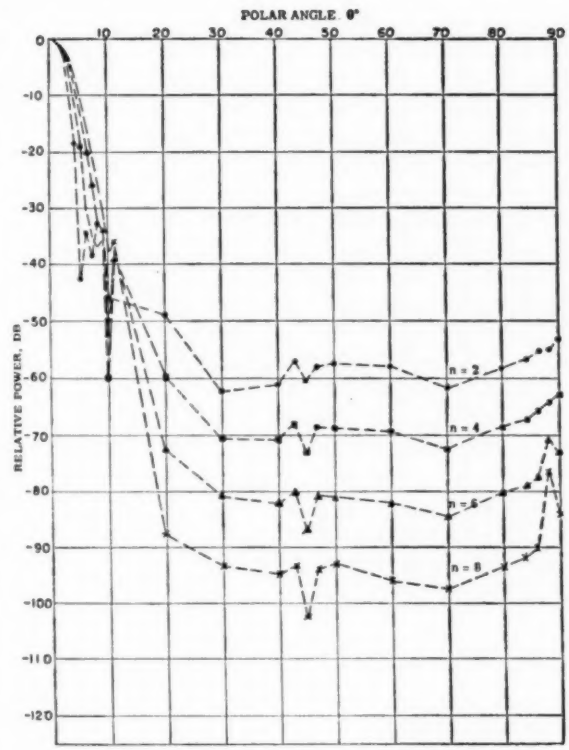
These patterns have been numerically calculated on IBM computers for a number of values of each of the parameters n, f/D, and D/λ, corresponding to primary



(a)



(c)



(b)

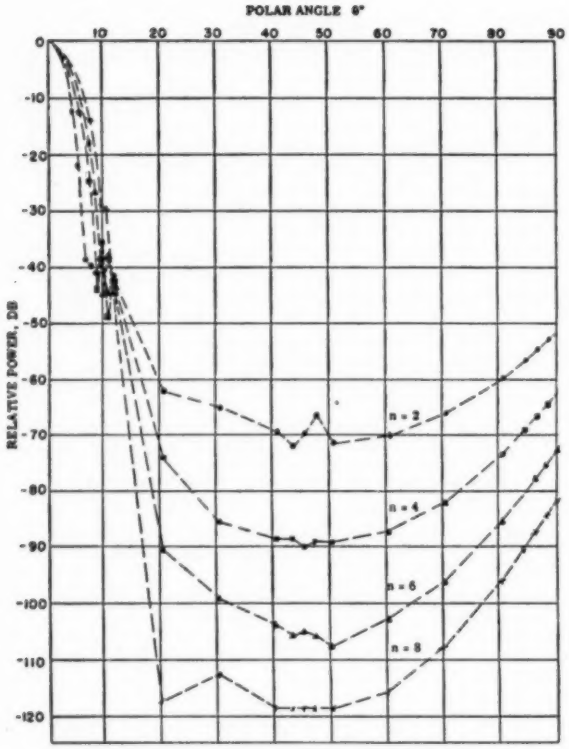
FIG. 4. (a) H-plane pattern, $D/\lambda = 25.4$. (b) H-plane pattern, $D/\lambda = 20.46$. (c) H-plane pattern, $D/\lambda = 10.6$. Diffraction patterns in principal H-plane for paraboloid reflector and primary feed gain function,

$$G_p = \begin{cases} 2(n+1)\cos^n\psi & \text{for } 0 \leqq \psi \leqq \frac{\pi}{2} \\ 0 & \text{for } \frac{\pi}{2} \leqq \psi \leqq \pi. \end{cases}$$

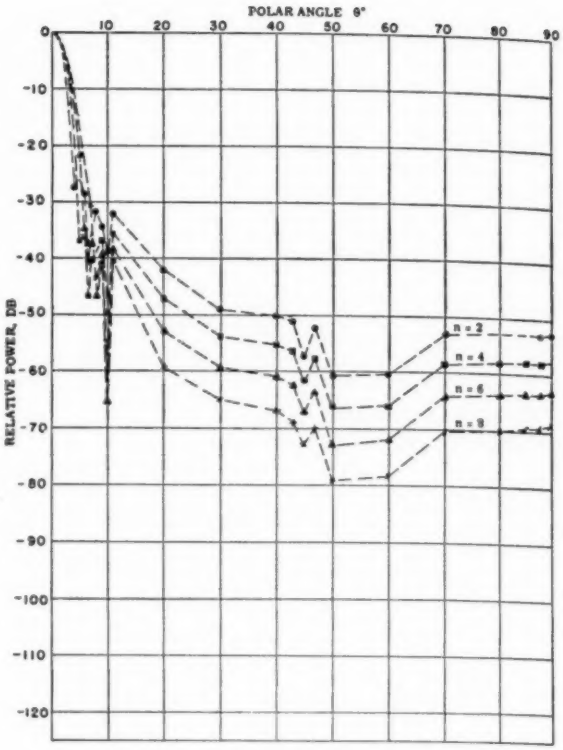
This grouping illustrates pattern changes due to changes in D/λ when f/D is held constant at 0.3. The E-plane patterns are similar in form and quantitative differences are not significant.

gain, aperture efficiency, and uniformly illuminated aperture gain, respectively. These have been plotted in Figs. 4 and 5.

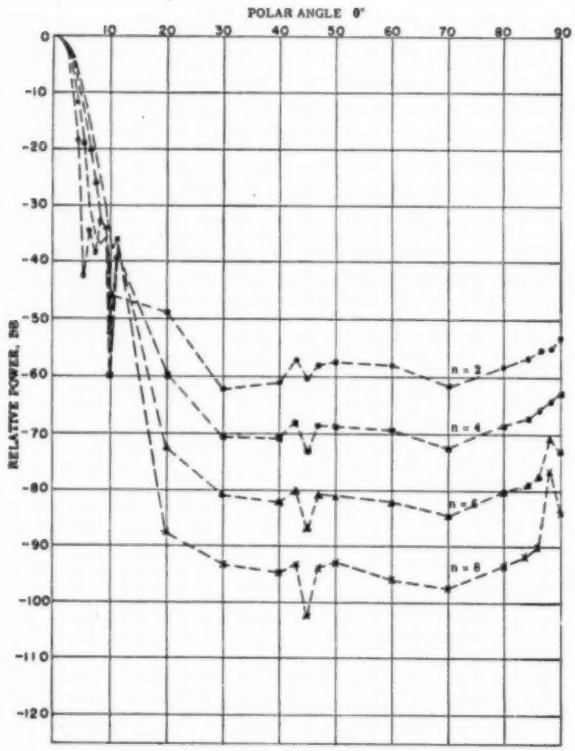
First it should be said that these patterns are not complete, since they were evaluated at discreet values of θ, and some lobe maxima and nulls were undoubtedly



(a)



(c)



(b)

FIG. 5. (a) H-plane pattern, $f/D=0.25$. (b) H-plane pattern, $f/D=0.3$. (c) H-plane pattern, $f/D=0.38$. Diffraction patterns in principal H-plane for paraboloid reflector and primary feed gain function,

$$G_p=\begin{cases}2(n+1)\cos^n\psi & \text{for } 0\leqq\psi\leqq\frac{\pi}{2}\\ 0 & \text{for } \frac{\pi}{2}\leqq\psi\leqq\pi.\end{cases}$$

This grouping illustrates pattern changes due to changes in f/D when D/λ is held constant at a value near 20.

missed. However, in the regions of most interest the intervals were made small in order to get the general level.

It can be seen that for small θ, all the patterns conform to the most significant results of aperture theory:

(a) As primary gain increases (n increases), secondary gain decreases (broader main beams).

(b) As primary gain increases (n increases), side lobe level decreases.

(c) As D/λ and f/D increase, so does secondary gain. (The variation with D/λ may be seen in Fig. 4. The f/D variation is indicated in Fig. 5.)

A significant departure from aperture theory is that minor lobe intensity does vary with D/λ. This is also true for the wide-angle energy.

In the region of greatest interest for this investigation the level at 90° was taken as an indication of the level of the diffracted energy at wide angles, and here the most important conclusions are as follows:

(a) The sharpest variation by far was with primary gain, going in the same direction as the small-angle side lobes. This may be seen in the variation of the 90° ordinates in each of Figs. 4(a) through 5(c). A typical variation is illustrated in Fig. 6, and it indicates roughly a drop of 7 db per db increase of primary feed gain.

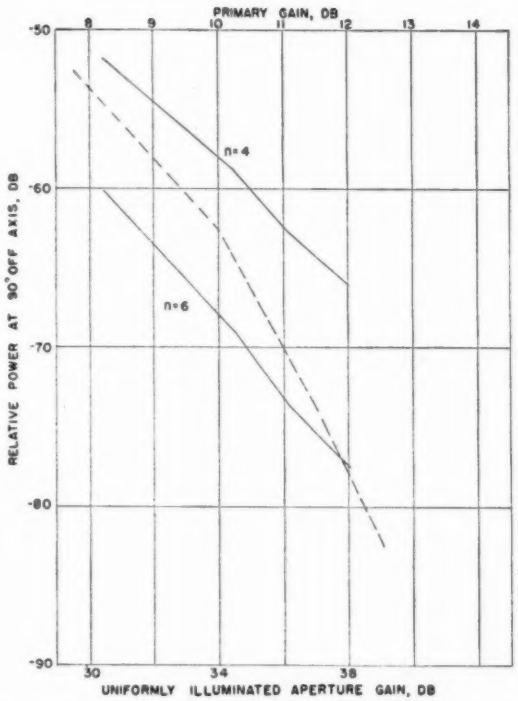


FIG. 6. Typical variation of wide-angle radiation level with primary feed gain and uniformly illuminated aperture gain. These curves give relative power at 90° off axis in the principal plane diffraction patterns for a paraboloid reflector of $f/D=0.3$ fed by a primary having a gain function,

$$G_p=\begin{cases}2(n+1)\cos^n\psi & \text{for } 0\leqq\psi\leqq\frac{\pi}{2}\\ 0 & \text{for } \frac{\pi}{2}\leqq\psi\leqq\pi.\end{cases}$$

The dashed line curve illustrates the variation with primary feed gain when $D/\lambda=20.4$. The two solid line curves illustrate the variation with uniformly illuminated aperture gain, utilizing the lower scale of abscissas.

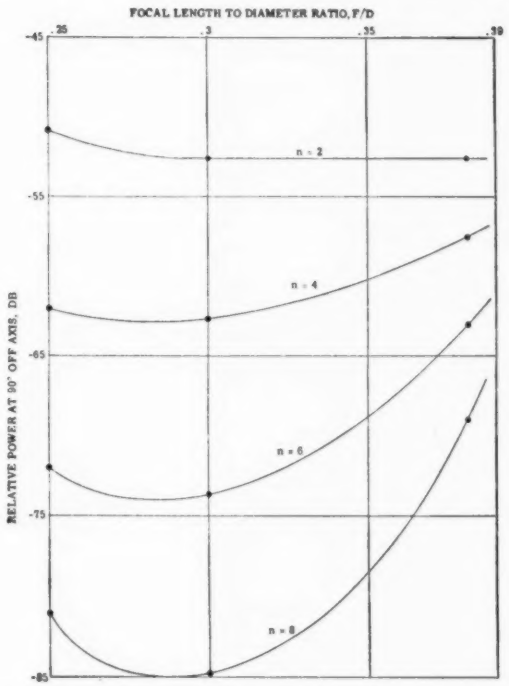


FIG. 7. Typical variation of wide-angle radiation level with focal length to diameter ratio. These curves give the relative power at 90° off axis in the principal plane diffraction patterns for a paraboloid reflector of $D/\lambda\approx 20$ fed by a primary having a gain function,

$$G_p=\begin{cases}2(n+1)\cos^n\psi & \text{for } 0\leqq\psi\leqq\frac{\pi}{2}\\ 0 & \text{for } \frac{\pi}{2}\leqq\psi\leqq\pi.\end{cases}$$

(b) The second significant factor was D/λ, the wide-angle radiation level decreasing with increasing D/λ. Thus, as shown in Fig. 6, for a tapered illumination corresponding to $n=6$, the 90° level decreased approximately 2 db for each db increase in the gain of the uniformly illuminated dish. The variation was smaller for smaller primary gains.

(c) The last factor, f/D, produced very small variations in the region of small f/D. As shown in Fig. 7, only when both f/D and primary gain are large does one obtain a sharp increase of wide-angle radiation level with increasing f/D.

(d) The H-plane and E-plane patterns differed very little. At wide angles the E-plane radiation level was always lower than the corresponding level in the H-plane, but this difference was never greater than 3 db over the range of parameters illustrated.

It should be noted that the requirements on primary gain are in opposite directions for the two contributions to wide-angle energy, i.e., reduction of the level due to directly radiated feed energy calling for a low gain feed whereas a high gain feed is required to reduce the level of the scattered energy. In conclusion, one might say, in general, that design for low side lobe at any angle

in the scattered pattern calls for large secondary and primary gain.

APPENDIX

Expanding the integral $\mathbf{I}$ in the coordinates of Fig. 3, for a paraboloidal reflector, it can be shown that

$$\mathbf{I}=\text{const}\int_0^{2\pi}\int_0^{\Psi} G_P^{\frac{1}{2}}\exp\{-jkf(\sec^2\psi/2)\times[1+\cos\theta\cos\psi-\sin\theta\sin\psi\cos(\xi-\phi)]\}\times[(\mathbf{N}\cdot\mathbf{e}_r)\mathbf{S}_1-(\mathbf{N}\cdot\mathbf{S}_1)\mathbf{e}_r]\frac{\tan\psi/2}{\cos\psi/2}d\psi d\xi.$$

There are several other simplifying assumptions which limit the accuracy of the results. One of these is that $\mathbf{S}_1=\mathbf{i}_z$ over the entire surface of the reflector. This implies that the far field of the primary feed is to wide-angle energy, i.e., reduction of the level due to more than just quasi-point source. This assumption follows from the previous assumptions applied to a paraboloid only if the feed is a true point source with a single center of phase. Furthermore, since the geometry of the paraboloid makes all reflected rays parallel to the axis, $\mathbf{e}_r$ can have no axial component and it is set equal to $\mathbf{i}_z$ at each point on the reflector. This assumes that the cross polarization component of $\mathbf{e}_r$ may be neglected in calculating the principal polarization diffraction pattern because the cross polarization energy is a very small fraction of the total energy in most feeds of interest. With these simplifications, the integral defining the scattered pattern becomes

$$\mathbf{I}=\text{const}\int_0^{2\pi}\int_0^{\Psi} G_P^{\frac{1}{2}}\exp\{-jkf(\sec^2\psi/2)\times[1+\cos\theta\cos\psi-\sin\theta\sin\psi\cos(\xi-\phi)]\}\times[\mathbf{i}_x+\mathbf{i}_z(\tan\psi/2)\cos\xi]\tan\psi/2\,d\psi d\xi,$$

and using the relations

$$\int_\alpha^{2\pi+\alpha}\exp(\pm ja\cos\xi)d\xi=2\pi J_0(a),$$

$$\int_\alpha^{2\pi+\alpha}\cos\xi\exp(ja\cos\xi)d\xi=2\pi jJ_1(a),$$

$$\int_\alpha^{2\pi+\alpha}\sin\xi\exp(ja\cos\xi)d\xi=0$$

to perform the integration over ξ one obtains the result quoted in the text.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 6 JUNE, 1955

Effect of Reactor Irradiation on the White-to-Grey Tin Transformation*

JEROME FLEEMAN† AND G. J. DIENES
Brookhaven National Laboratory, Upton, New York

(Received October 21, 1954)

The effect of low-temperature (liquid nitrogen) reactor irradiation on the white-to-grey tin transformation has been investigated. It was found that, compared to an unirradiated pure sample, the transformation, measured by dilatometry, is drastically accelerated by prior irradiation. The irradiation apparently eliminates the normally very long induction period. The kinetic behavior of reactor irradiated samples and of samples seeded with grey tin were found to be qualitatively similar. These results indicate that the defects introduced into white tin by reactor irradiation serve as nuclei, or at least embryos of nucleation, for the subsequent phase transformation. Whether point defects, their aggregates, or the strains surrounding the displaced atoms are responsible for the artificial "seeding" cannot yet be decided. Reactor irradiation appears to be a promising new tool for studying nucleation and growth processes since quite uniform artificial nucleation can be accomplished this way.

I. INTRODUCTION

IT is well known that reactor irradiation results in large changes in the physical properties of solids.[1] These phenomena are generally interpreted in terms of lattice defects and their aggregates produced by the knocking about of the atoms of the crystal lattice. Displaced atoms, or strained regions of the crystal around defects (particularly interstitials), may serve as nuclei for solid state reactions. Murray and Taylor[2] investigated the effect of neutron irradiation on a supersaturated solid solution of beryllium in copper. From a study of various physical property changes, they concluded that the evidence pointed toward the

* Work carried out under contract with the U. S. Atomic Energy Commission.

† Deceased September 30, 1954.

[1] J. C. Slater, J. Appl. Phys. **22**, 237 (1951); G. J. Dienes, Ann. Rev. Nuclear Sci. **2**, 187 (1953), and J. Appl. Phys. **24**, 666 (1953).

[2] G. T. Murray and W. E. Taylor, Acta Metallurgica **2**, 52 (1954).

formation of precipitate nuclei during irradiation. The interpretation of the results in a system as complex as a precipitation alloy relies of necessity on rather indirect evidence.

The white-to-grey tin transformation offers many advantages for a study of the effects of reactor irradiation on phase transformation. These advantages are[3]: (a) the transformation is a first-order phase change between two well-known structures in a monatomic system, (b) the transformation temperature 13.2°C is in a convenient temperature range, (c) the transformation rate is slow and becomes negligible below about −80°C, (d) the radioactivity produced in the reactor in tin causes no appreciable experimental difficulties, and (e) the transformation is easily followed by dilatometry by virtue of the very large volume increase accompanying the white-to-grey tin phase change (27 percent). An experimental study of the white-to-grey tin transformation subsequent to low-temperature irradiation of the white tin samples is described in this paper.

II. EXPERIMENTAL

Electrolytically refined tin (99.995 percent) obtained from the Vulcan Detinning Company was used in these experiments. A set of cylindrical samples of about $\frac{1}{2}$ cm³ (approximately $\frac{1}{4}$ in. diameter and $\frac{5}{8}$ in. long, annealed in vacuum for 1 hr at 100°C) was irradiated in the low-temperature facility[4] at the Brookhaven reactor. The irradiation was carried out at liquid nitrogen temperature to an integrated neutron flux of about 10^{18} nvt. The samples were accurately weighed prior to irradiation. After removal from the reactor the specimens were stored in liquid nitrogen while the induced radioactivity decayed. The specimens were transferred directly from liquid nitrogen to dilatometers at controlled low temperatures (−60 to −20°C range) for transformation rate studies. Control specimens of unirradiated tin had the same thermal history as the irradiated ones.

The dilatometers were of conventional design using ethyl alcohol as the dilatometer fluid. The unirradiated samples showed no transformation during the kinetic experiments and fluctuations in the dilatometers containing them were used as corrections for temperature fluctuations. In every case these corrections were minor. The over-all estimated precision is 0.5 percent over the complete transformation range in any given experiment.

A typical rate curve for irradiated tin is shown in Fig. 1 where the percent change in volume is plotted as a function of time at reaction temperature. The change in volume, ΔV, is $V - V_i$ where V is the volume at any time t and V_i is the initial volume of the pure white tin calculated from the weight of the sample. According to Fig. 1 the transformation process in irradiated tin is characterized by a rather short induction period (defined as the intercept of the straight line on the time axis in Fig. 1) followed by a linear volume *vs* time region and a final very slow region where the transformation is essentially complete. The transformation rate, as measured by the volume change, was found to be constant in every case at least from 2 to 20 percent on the $\Delta V/V_i$ scale (i.e., $\Delta V/V_i$ linear with t over practically the whole transformation range). Unirradiated tin during such an experiment shows no transformation at all and would be represented by the abscissa itself. As a matter of fact, the pure unirradiated tin samples were stable at the reaction temperature of −20°C to −60°C for at least one week and probably much longer. Appreciable transformation in the absence of radiation could only be induced in these samples by seeding with grey tin or by heavy cold work at low temperature.

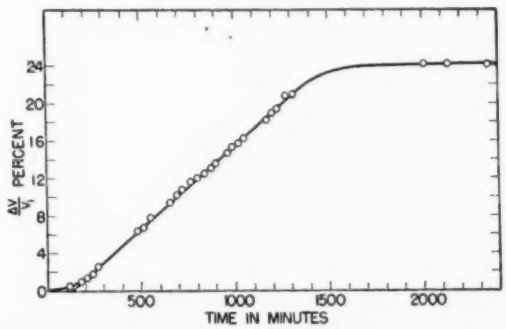


FIG. 1. Percent volume change *vs* time curve for the white-to-grey tin transformation following reactor irradiation of the white tin at liquid nitrogen. Temperature of transformation reaction = −50.3°C. Annealed pure white tin shows no transformation and would be represented by the abscissa in the figure.

Concurrently with the experiments on irradiated tin rate studies were made on tin seeded with grey tin. The seeding was done by simply touching one face of the sample to grey tin at the reaction temperature. The volume *vs* time curves for the seeded samples were identical in shape with those for the irradiated ones. However, the data for the seeded samples were found to be far less reproducible than for the irradiated ones. The poor reproducibility of transformation rates for all types of unirradiated tin has been noted in the literature.[3]

The results are given in Table I where the induction periods and the constant rates are listed for irradiated and seeded tin at various reaction temperatures. The data show clearly that the steady rates are quite reproducible in the case of the irradiated tin while the scatter is much greater in the case of the seeded samples. The transformation rates are of similar magnitude. The reproducibility of the induction periods is seen to be quite poor.

As far as the effect of reaction temperature is concerned, a maximum in the steady rate of transformation

[3] For recent work on the tin transformation see: F. W. Cagle, Jr., and H. Eyring, J. Phys. Chem. **57**, 942 (1953), and R. R. Rogers and J. F. Fydell, J. Electrochem. Soc. **100**, 383 (1953), where references to the earlier literature are given.

[4] McReynolds, Augustyniak, McKeown, and Rosenblatt, Phys. Rev. **94**, 1417 (1954).

TABLE I. Kinetic data for white-to-grey tin transformation.

Reaction temperature, °C	Irradiated white tin[a] Induction period, minutes	Irradiated white tin[a] Rate of transformation in percent volume change per minute	White tin seeded with grey tin[a] Induction period, minutes	White tin seeded with grey tin[a] Rate of transformation in percent volume change per minute
−59.5	500	21.6×10^{-3}	400	17.8×10^{-3}
	500	19.5×10^{-3}		
−50.3	150	17.8×10^{-3}	300	28×10^{-3}
	500	13.0×10^{-3}	500	36×10^{-3}
−40.7	100	30×10^{-3}	80	116×10^{-3}
	120	30×10^{-3}	210	67×10^{-3}
−30.0	90	40×10^{-3}	75	90×10^{-3}
	150	42×10^{-3}	210	45×10^{-3}
−20	400	23×10^{-3}	No measurable transformation in 2000 minutes.	
	530	21.5×10^{-3}		

[a] Pure annealed and untreated white tin samples, which served as controls, were stable at these temperatures (no transformation) for at least one week and probably much longer.

occurs near −30°C. This is in agreement with data in the literature. Tammann and Dreyer[5] investigated the growth of pustules of grey tin on a white tin surface at various temperatures. They observed that the radius of the pustule increased linearly with time at a fixed temperature with the pustule spreading from the point of innoculation. They give the temperature of maximum growth rate as −30°C. It should be noted that according to their results, and the data presented here, transformation on the surface as well as volume transformation occurs at a constant rate.

X-ray diffraction patterns taken at liquid nitrogen temperature after a liquid nitrogen irradiation proved quite conclusively that no significant transformation occurred during the irradiation itself since no grey tin lines were detected. The upper limit on the amount of grey tin possibly present but undetectable in the diffraction patterns was calculated as 0.05 percent. In the rate studies, therefore, the starting material was "seeded" but untransformed white tin.

The theoretical 27 percent change in volume was not obtained in any of the kinetic runs. The transformation apparently stopped at $\Delta V/V_i$ values between 24 and 25 percent. In the final state of transformation the material is a powder. No observations were made on the character of the transformation during the rate studies, but the sample certainly crumbled before the end of the run. There may be, therefore, considerable inaccuracies in the volume measurements for the last few points. If the strong retardation at high degree of conversion is real the explanation is probably that of crystallographic interference. More detailed studies will be required to settle this point.

[5] G. Tammann and K. L. Dreyer, Z. anorg. u allgem. Chem. **199**, 97 (1931).

III. DISCUSSION

These experiments show clearly that low-temperature neutron irradiation has a drastic effect on the white-to-grey tin transformation. Compared to a pure untreated sample the transformation is tremendously accelerated apparently by an essentially complete elimination of a normally very long induction period.

The over-all features of the transformation after irradiation and after seeding with grey tin are similar. In both cases the volume rate of transformation was found to be constant over practically the whole transformation. The higher degree of reproducibility of the kinetic data for the irradiated specimens indicates that the irradiation resulted in quite uniform "seeding" of the white tin.

The foregoing results indicate that the defects introduced into white tin by reactor irradiation serve as nuclei, or at least embryos of nucleation, for the subsequent phase transformation. Whether point defects, their aggregates, or the strains surrounding the displaced atoms are responsible for the seeding cannot yet be decided. Reactor irradiation, or the formation of displaced atoms in general by fast particle irradiation, appears to be a promising new tool for studying nucleation and growth processes since quite uniform artificial nucleation evidently can be accomplished this way.

The kinetic data are not yet extensive enough to permit a detailed analysis. The constancy of the volume rate of transformation at a fixed temperature is somewhat puzzling. The simplest interpretation is that nucleation is completed by the irradiation and that the material transforms by unidimensional growth (thin rods). Other mechanisms based on nucleation and growth could also be postulated. We feel that much further data will have to be obtained, particularly x-ray data on the crystallography of the transformation, before the kinetic picture can be clarified.

Interpretation of the temperature dependence of the transformation rate is closely related to and depends on the interpretation of the volume rate-time data. Suffice it to say here that the over-all behavior can be explained either on nucleation and growth basis or by simple rate theory as discussed by Cagle and Eyring.[3] We feel that the final answer will depend on a quantitative analysis of extensive kinetic data again in conjunction with detailed crystallographic information. Further work is planned along these lines.

ACKNOWLEDGMENTS

The authors are grateful to R. Smoluchowski, G. H. Vineyard, D. Keating, and Paul W. Levy for stimulating and critical discussions and to D. Keating for carrying out the x-ray measurements.

Monte Carlo Calculation of Gamma-Ray Albedos of Concrete and Aluminum*

J. F. PERKINS†
Convair, Fort Worth, Texas
(Received September 3, 1954)

Number and energy gamma-ray albedos have been calculated for a material of $Z_{eff}=13$, corresponding to both concrete and aluminum. The Monte Carlo method was used, the calculations being performed on an IBM-701. The single- and multiple-scattered components were evaluated separately; the multiple-scattered component was considerably the larger of the two in most cases, being as much as three times as large as the single-scattered component. For moderately small incident angles, the emergent photons are distributed approximately proportional to the cosine of the normal angle of emergence. The spectra peak around 150–250 kev and in some cases have a second peak at higher energy; such secondary peaks move upward in energy and increase in importance as the angle of incidence increases. The average energy of the emergent photons varies from 0.33 to 1.4 m_0c^2. For the multiple-scattered component the average energy varies from 0.30 to 0.93 m_0c^2. Buildup factors have been calculated for 2 m_0c^2 gammas normally incident on aluminum of 2 and 4 mean free paths thickness.

HAYWARD and Hubbell[1,2] have recently calculated the number and energy albedos, i.e., reflection coefficients, of several materials for 1-Mev photons. They used a Monte Carlo technique, calculating the case histories by hand. In spite of the tediousness of such calculations (each case history required about 1 man-day to calculate) they were able to extract an impressive amount of information from the 67 case histories which they treated.

The calculations described here deal with the albedos of concrete, for which a typical $Z_{eff}^4=\langle Z^5\rangle/\langle Z\rangle=13.5^4$. For convenience a value of $Z=13$ was used in these calculations; hence these results apply also to aluminum. The insignificance of the error arising from this choice of Z follows from the very slow variation of albedos with Z.[1,2]

A variety of incident energies and angles have been treated and information has been obtained concerning the distribution of the emergent radiation in direction and energy. A variation of the straight analog method was employed in order to reduce the variance in the calculated values of the emergent distribution. The necessary sample size has dictated the use of a high-speed computer; the problem was coded and run on an IBM-701.

The photon paths were followed in a straightforward manner except as noted in the following; the formulas for compounding of angles, etc., have been conveniently set forth elsewhere.[2] The uniformly distributed random numbers N used in the calculations were generated as required, using the method of middle squares. The Compton scattering angle Θ was chosen by the rejection method. The quantity Θ was first chosen from the distribution $f(\Theta)=A\cdot\sin\Theta$ by putting $\cos\Theta=2N_1-1$. Then $(d\sigma/d\Theta)/\sin\Theta$, which is a rational function of energy and of $\cos\Theta$, was evaluated. If $N_2<(1/r_0^2)\times(d\sigma/d\Omega)$ this choice of Θ was accepted; if the inequality did not hold a new N_1 and N_2 were generated and the process repeated until the inequality was satisfied.

Only Compton scattering and photoelectric absorption processes were considered. Scattering cross sections were calculated from the Klein-Nishina formula, while photoelectric cross sections were evaluated by interpolation in a table taken from results of Davisson and Evans[3] and White.[4]

A survival factor S was associated with each segment of a photon's path. Having chosen a point of interaction, the new direction of the photon was calculated on the assumption that it was indeed scattered rather than absorbed. The probability P that the photon escaped the surface was then calculated from the new energy α according to the equation

$$P=e^{-\left|\frac{\sigma_t(\alpha)Z}{\cos\theta}\right|} \quad \text{if } \cos\theta<0$$

$$P=0 \quad \text{if } \cos\theta\geq 0.$$

This was in keeping with the general rule that the portions of a calculation which are easy to treat deterministically should be so treated, and only the difficult parts should be treated by sampling methods.[5] The estimates of the number and energy albedos obtained from the path of a single photon are

$$\sum_i P_i S_i \quad \text{and} \quad \sum_i(\alpha_i P_i S_i)/\alpha_0.$$

The choice of the distance to the point of the next interaction was biased so as to assure that this point would lie within the medium. This arrangement provided for including a contribution from each interaction, thus reducing the variance in the results, particularly in the tabulation of the emergent distributions. After the ith scattering the new value of S

* Work performed under contract with the Wright Air Development Center of the U. S. Air Force.
† Now at Lockheed Aircraft Corporation, Marietta, Georgia.
[1] E. Hayward and J. Hubbell, Phys. Rev. **93**, 955 (1954).
[2] E. Hayward and J. H. Hubbell, Natl. Bur. Standards Rept. NBS 2768 (1953).
[3] C. M. Davisson and R. D. Evans, Revs. Modern Phys. **24**, 79 (1952).
[4] G. R. White, Natl. Bur. Standards Rept. NBS 1003 (1952).
[5] H. Kahn, Nucleonics **6**, 27 (1950).

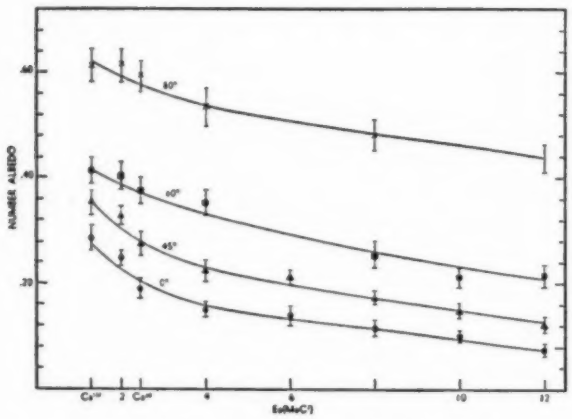


FIG. 1. Total number albedos for photons incident at $\theta_0=0°$, 45°, 60°, and 80° as a function of incident energy.

becomes $S_{i+1}=S_i(\sigma_{kn}/\sigma_t)(1-P_i)$. The second term gives the probability that the interaction was a scattering rather than an absorption, while the last term corrects for the bias introduced by forcing the photon to remain with the medium. Each photon was followed until S became less than 0.05.

The contribution to the albedos arising from the first scattering can be calculated by a straightforward deterministic process, and it is the contribution from subsequent scatterings which requires a Monte Carlo treatment. For this reason these two contributions to the albedos were calculated separately.

The rather large sample size required for moderate accuracy necessitated a computer program which would require a minimum amount of calculating time. For this reason the programing was done in "fixed point," rather than by use of an abstract interpretive system. This necessitated a considerable amount of tracing, i.e., printing out and examining of the step-by-step performance of the calculation, to assure that there were no scaling errors over the range of energies and angles considered. The program is easily contained within the high-speed electrostatic storage of the 701.

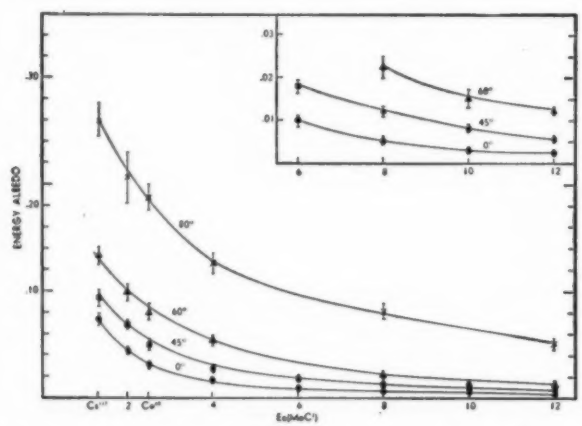


FIG. 2. Total energy albedos for photons incident at $\theta_0=0°$, 45°, 60°, and 80° as a function of incident energy.

Approximately 80 percent of the machine time is required by the subroutines which calculate $\sin x$, e^x, $\log x$, and $\sqrt{x}$ by iterations. This leaves some latitude for speeding up the program for later use by means of evaluation of these functions by table lookups. As presently programed a sample of 250 gamma-ray case histories requires about 8 minutes of calculating time and about $1\frac{1}{2}$ minutes of input-output time.

The calculations have dealt with four incident angles θ_0 and a variety of incident energies, including those of Cs^{137} and Co^{60}. The values of the total albedos are shown in Figs. 1 and 2. Figure 3 shows the energy albedos for normal incidence broken down into the single- and multiple-scattered components; the ratio of multiple- to single-scattered component is even larger in the case of the number albedos. The single-scattered component is the smaller of the two in all the cases treated here except the energy albedos for $\theta_0=80°$.

The standard deviations indicated in the figures

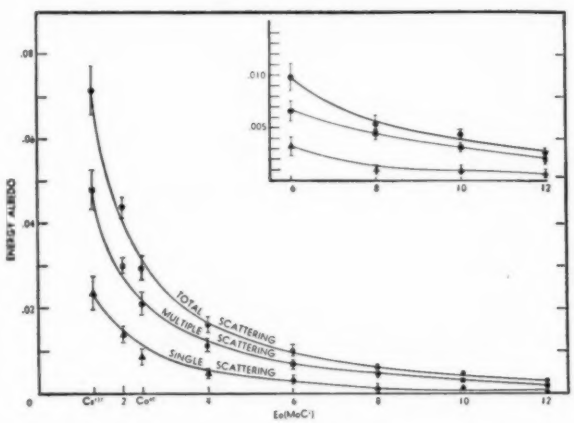


FIG. 3. Total energy albedos and the single- and multiple-scattered components of this albedo for $\theta_0=0°$, as a function of incident energy.

were calculated according to the equation $\sigma=[(\langle X^2\rangle-\langle X\rangle^2)/N]^{\frac{1}{2}}$, where X represents PS and $\alpha PS/\alpha_0$ for the number and energy albedos, respectively. Calculations were performed in blocks of $N=250$ incident photons, a sample size which resulted in fractional standard deviations ranging from ± 5 percent to ± 20 percent. For most of the cases only one set of 250 photons was treated. For 6 combinations of incident angles and energies, for each of which there are 4 calculated albedos (with some correlations between the 4), two such sets were treated using different initial random numbers. A statistical test was made on these to determine whether the computed σ's appropriately represent the statistical uncertainty. The equation $Y=(X_1-X_2)/(\sigma_1^2+\sigma_2^2)^{\frac{1}{2}}$ was calculated for these 24 albedos, with the following results: $\langle Y^2\rangle=0.98$, $\langle|Y|\rangle=0.88$, $\langle Y\rangle=0.511$. These results seem to be in satisfactory agreement with the assumption that Y is normally distributed with mean 0 and standard

deviation 1, indicating that the computed σ's do indeed accurately represent the statistical uncertainty.

Six of the albedos which have been calculated here are directly comparable with results of Hayward and Hubbell;[1,2] the agreement between the two calculations is satisfactory. The energy albedos fall off with increasing incident energy much more rapidly than do the number albedos because of the greater fractional energy degradation at the higher energies. The average energy of the emergent photons varies from about 0.33 m_0c^2 for normally incident Cs^{137} radiation to 1.4 m_0c^2 for 12 m_0c^2 radiation incident at $\theta_0=80°$. The corresponding range for the multiple-scattered component alone is 0.30 m_0c^2 to 0.93 m_0c^2.

The emergent energy spectra show a peak in the region of 150–250 kev, the position of which does not vary greatly with incident angle or energy. There is also a smaller peak at a higher energy, the position of which increases with increasing incident energy. This behavior of the secondary peak is in accord with experimental results of Hayward and Hubbell.[6,7] Hine

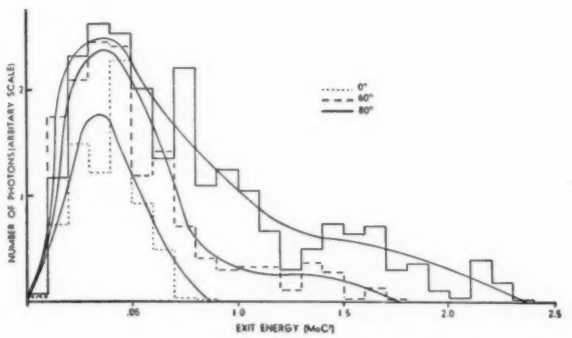


FIG. 4. Energy spectra of emergent photons for Co^{60} gamma rays incident at $\theta_0=0°$, 60°, and 80°.

and McCall[8] have also noted the presence of a secondary peak. These experimental results did not come to our attention until the present calculations had been performed and exploration of the behavior of this peak has not been emphasized. We note that the relative size of the secondary peak increases with increasing incident angle. Figure 4 shows energy spectra for a Co^{60} radiation source.

Figure 5 shows the distribution of the number and energy of emergent photons over the polar angle of emergence θ integrated over the azimuthal angle of emergence ϕ for the case of normally incident Co^{60} radiation. For this particular case the distributions are strongly anisotropic and both the multiple-scattered component and the total dose have approximately the form $\cos\theta$. For large angles of incidence the multiple-scattered component becomes slightly more nearly

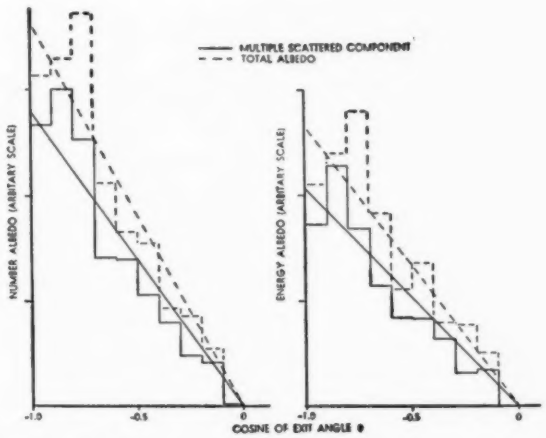


FIG. 5. Variation of number and energy of emergent photons as a function of polar angle of emergence for Co^{60} gamma rays incident at $\theta_0=0°$.

isotropic and the single-scattered component, which is always fairly isotropic, becomes relatively more important.

Figure 6 shows the distribution over the azimuthal angle ϕ with the θ dependence integrated out. As would

TABLE I. Number and energy buildup factors for $2m_0c^2$ gamma rays normally incident on aluminum slabs (present results) and for Cs^{137} gamma rays normally incident on water slabs (Berger's results).

Slab thickness in mean free paths	2 m_0c^2 gammas normally incident on aluminum		Cs^{137} gammas normally incident on water	
	B_N	B_E	B_N	B_E
2	3.12±0.14	2.00±0.08	3.93	2.50
4	5.26±0.69	2.80±0.30	8.2	3.66

be expected, the forward ($\phi=0$) peaking is stronger for the larger incident angles and arises largely from the single-scattered component.

As an incidental part of this work number and energy buildup factors, B_N and B_E, were calculated for 2 m_0c^2 gammas normally incident on aluminum slabs of 2

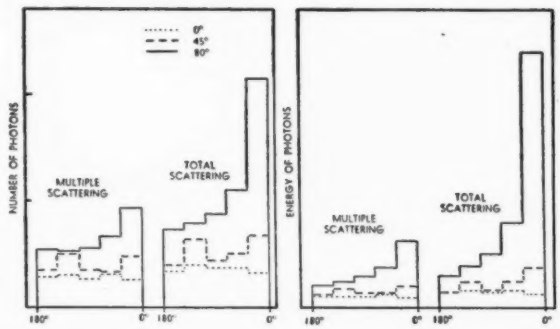


FIG. 6. Variation of number and energy of emergent photons as a function of azimuthal angle of emergence for Co^{60} gamma rays incident at $\theta_0=0°$, 45°, and 80°. The multiple-scattered component and the total dose are shown separately.

[6] E. Hayward and J. H. Hubbell, J. Appl. Phys. **25**, 506 (1954).
[7] E. Hayward and J. H. Hubbell, Natl. Bur. Standards Rept. NBS 2264 (1953).
[8] G. J. Hine and R. C. McCall, Nucleonics **12**, No. 4, 27 (1954).

and 4 mean free paths thickness. These results are shown in Table I. The sample used in each of these two cases also consisted of 250 incident photons. These particular cases do not seem to have been treated before, but Goldstein *et al.*[9] and Berger[10] have treated fairly similar cases. In contrast with Berger and the present author, Goldstein calculated the dose received by an isotropic detector, and hence weighted emergent photons with the factor $\sec\theta$. For this reason Goldstein's values are somewhat higher than the present ones. The present results seem to be in reasonable agreement with Berger's results for Cs^{137} gammas incident on water. Berger's results were obtained from a Monte Carlo computation performed by hand. They dealt with lower-energy gammas and a higher-Z material than treated here; hence his buildup factors would be expected to be somewhat larger.

[9] Goldstein, Wilkins, and Presier, NDA Memo 15C-20 (1953).
[10] M. J. Berger, results presented at the Symposium on Monte Carlo Methods, Gainesville, Florida, March 16–17, 1954.

ACKNOWLEDGMENTS

The author would like to express his appreciation to Mr. F. W. Donaldson for assistance with the programing.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 6 JUNE, 1955

Reverse Current and Carrier Lifetime as a Function of Temperature in Germanium Junction Diodes

E. M. PELL
General Electric Research Laboratory, Schnectady, New York
(Received November 26, 1954)

The reverse current and carrier lifetime have been measured in a series of germanium diodes as a function of temperature between room temperature and liquid nitrogen temperature. The lifetime reaches a plateau at low temperatures, and its behavior can be explained in terms of the Hall-Shockley-Read recombination theory. Log i_R *vs* $1/T$ exhibits a break to a shallower slope at lower temperatures, which can be explained in terms of charge generation by recombination centers in the space-charge region.

I. INTRODUCTION

THE Hall-Shockley-Read theory of recombination in germanium[1,2] predicts that the plot of lifetime *vs* reciprocal temperature should reach a plateau at low temperatures. The presence of this plateau has never been adequately verified experimentally.[3] The present experiment demonstrates that such a plateau does in fact exist and offers some evidence for a model which would explain the large variations in slope which are observed on such a plot at intermediate temperatures.

Since the lifetime was measured by making a junction diode with the germanium under study, it was convenient also to measure the diode reverse current as a function of temperature. On a plot of log i_R *vs* $1/T$, two straight-line regions were evident in all diodes studied, and the slopes of these regions were the same for all samples to within experimental error. It was found that the steep region fitted the usual junction diode theory[4] in both slope and magnitude. The shallower region could be fitted, in slope and magnitude, using a model of charge generation by recombination centers in the space-charge region. This model is a necessary outgrowth of the Hall-Shockley-Read recombination theory and has already been alluded to in explaining observed data on silicon diodes.[5] This paper examines this model more fully and shows how this phenomenon can be used for finding the position of the recombination (or, in this case, generation) centers.

II. EXPERIMENTAL TECHNIQUES

The germanium samples studied were in the form of diodes, either grown junctions or indium dots on N-type germanium. They are described in Table I, and it can be seen that in addition to representing two different methods of fabrication, they represent diodes with both N- and P-high-resistivity regions, and that there is a wide range of resistivities and perimeter-to-area ratios. All diodes were electrolytically etched in sodium hydroxide and mounted in a glass-and-metal vacuum system for the measurements. The system was evacuated by a mercury diffusion pump, and the gasket seal was of metal. Heat shields were employed around the sample, and the system was made light-tight (photo currents from residual light leakage were too small to read).

The temperature of the sample was measured with a calibrated chromel-constantan thermocouple attached to a copper plate to which the sample diode was soldered (with solder of 58°C melting point). An

[1] R. N. Hall, Phys. Rev. **87**, 387 (1952).
[2] W. Shockley and W. T. Read, Jr., Phys. Rev. **87**, 835 (1952).
[3] Some experimental information is given in the first two references. See also D. Navon and H. Y. Fan, Phys. Rev. **83**, 911(A) (1954), and Burton, Hull, Morin, and Severiens, J. Phys. Chem. **57**, 853 (1953); also P. Ransom and F. W. G. Rose, Proc. Phys. Soc. (London) **B67**, 646 (1954).
[4] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950).
[5] K. G. McKay and K. B. McAfee, Phys. Rev. **91**, 1079 (1953).

additional experiment with a sample crystal and a second thermocouple established that there was a negligible temperature difference between sample and base plate over the temperature range used.

Changing the sample geometry established that there were no observable effects to be ascribed to thermal stresses produced by the method of mounting.

Reverse current was measured with a breaker amplifier of 100 K ohms input impedance, arranged with suitable shunts to measure current, and followed by a recorder. Contact potentials in this instrument and the rest of the circuit were low enough at all sample temperatures so that good data could be obtained with only 0.001 volt across the diode.

Recombination rates were measured with a pulse method[6,7] using the diode itself to inject the minority carriers and to measure the decrease in their density as a function of time following injection. This method has been compared with other methods on six samples—both short and long lifetime—at room temperature, and the agreement was within 20 percent. The actual technique used was to compare an oscilloscope display of the reverse-current decay after pulsing with a family of curves, calculated from theory, using a half-silvered mirror to bring the planes of the two images into coincidence. The system was so designed that the sweep speed, which could be read from the dials on the oscilloscope, gave the lifetime of the minority carriers directly, after it was adjusted to bring the two curves into coincidence. This technique also made it possible to include the scope distortion in the coordinate system on which the theoretical curves were plotted, so that the entire area of the cathode-ray tube could be used. By the above method, the lifetime could be read in a fraction of a minute to within a few percent.

Trapping effects (generally trapping of holes[8]) have never been very serious in this method of measuring recombination rates. In an N-type sample, hole traps would not be expected to be bothersome since an experiment which measures minority carrier density would observe the trapped holes only when they re-evaporate. In a P-type sample, the situation would be expected to be worse since one trapped hole would give rise to a continuous current of electrons. Trapping effects were sometimes observed in diodes whose P-region was of higher resistivity than the N-region, but a small amount of ambient light always permitted τ to be read.

TABLE I. Ratio of observed magnitude of reverse current to magnitude calculated using previously published intrinsic resistivity data for estimate of partition functions. The "diffusion component" is $N_v N_c/[3.1\times10^{32}T^3\exp(c/k)]$. The "charge generation component" is $[N_v N_c/3.1\times10^{32}T^3\exp(c/k)]^{\frac{1}{2}}$.

Sample	Diffusion component	Chg. gen. component
1. In dot on 24Ω cm N, 0.01 cm²	0.92	<1.1
2. Grown, 4.2Ω cm N −0.4Ω cm P, 1.16 cm²	2.86	1.1
3. Grown, 4.2Ω cm N −0.4Ω cm P, 0.048 cm²	4.54	3.2
4. Grown, 0.4Ω cm P −0.06Ω cm N, 0.992 cm²	3.62	≥0.72
5. Grown, 0.4Ω cm P −0.06Ω cm N, 0.0921 cm²	6.05	<0.53
6. Grown, 12Ω cm P −4.4Ω cm N, 1.016 cm²	2.64	≥0.56
7. Grown, 12Ω cm P −4.4Ω cm N, 0.047 cm²	4.59	≥0.58
8. In dot on Fe doped Ge, 36Ω cm N, 0.132 cm²	0.22	1.94
9. Grown, 13Ω cm N −0.8Ω cm P, 0.028 cm²	1.62	1.20
10. In dot on Au doped Ge, 14Ω cm N, 0.0252 cm²	0.60	1.38

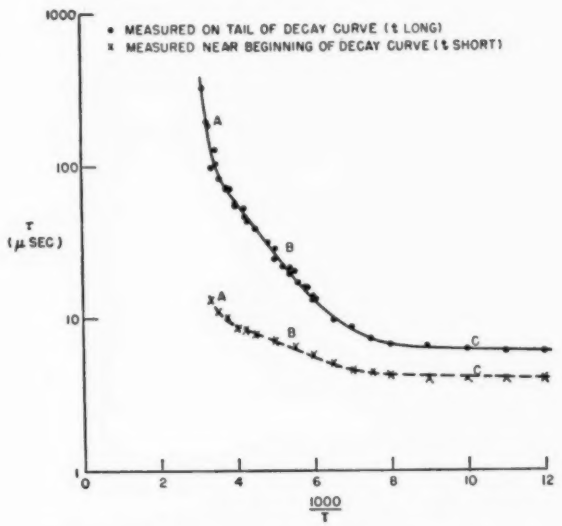


FIG. 1. Log carrier lifetime *vs* $1000/T$, grown junction (doped): 4.2Ω cm N *vs* 0.4Ω cm P.

III. SURFACE EFFECTS

Because the impedance of a reverse-biased junction can be very high, any surface phenomenon which could give rise to leakage or allied effects could be very troublesome. Several attempts were made to eliminate the possibility of surface effects. The perimeter-to-area ratio of the junctions was varied by cutting a single junction into two pieces with area ratios of about 20 to 1, but this method was too insensitive to be conclusive. A second attempt consisted of deliberately mistreating the surface by the introduction of various kinds of gas, with a spark discharge near the diode. The lifetime was not affected, but the reverse current was. The insensitivity of the lifetime to surface conditions, plus its excellent reproducibility, causes us to feel that the lifetime measurements were free of any appreciable errors caused by surface recombination. In the case of the reverse current, there is more room for doubt. To minimize any surface effect that might be present, reverse currents were read at as low a voltage as possible, because many recognizable surface effects have in our experience exhibited current increases with voltage. Currents were read at voltages down to 0.001 volt, using the theoretical diode characteristic corresponding to the measured diode temperature to find the saturation current. This generally gave a somewhat lower current than would have been obtained at higher

[6] E. M. Pell, Phys. Rev. **90**, 278 (1953).
[7] B. Lax and S. F. Neustadter, J. Appl. Phys. **25**, 1148 (1954). This is a more complete analysis of the method though all our measurements were made within the range of validity of the approximations in the previous reference.
[8] J. R. Haynes and J. A. Hornbeck, Phys. Rev. **90**, 152 (1953).

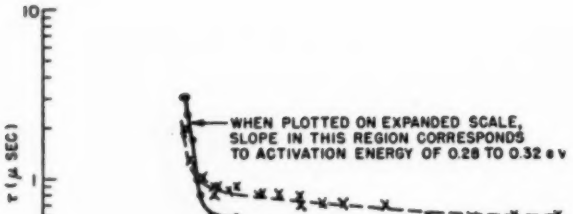

WHEN PLOTTED ON EXPANDED SCALE, SLOPE IN THIS REGION CORRESPONDS TO ACTIVATION ENERGY OF 0.28 TO 0.32 e v
<0.6
• MEASURED ON TAIL OF DECAY CURVE
× MEASURED NEAR BEGINNING OF DECAY CURVE
τ (μ SEC)
1000/T

FIG. 2. Log carrier lifetime *vs* $1000/T$, In dot junction on iron-doped germanium (36Ω cm *N*).

voltages, but using higher and more conventional voltages would not have seriously altered the resulting characteristic. In all units, there was a lower limit to the reverse current at each temperature which could not be improved upon with further etching, and the slope of this lower limit *vs* temperature was the same for all units tested. This implies that the measured characteristic was either the desired bulk property or was a reproducible surface effect. We lean toward the former interpretation because the magnitude of the current was at all temperatures so close to the value predicted from the bulk properties.

IV. RESULTS AND DISCUSSION

Recombination Rate *vs* Reciprocal Temperature

Figure 1 is a plot of carrier lifetime *vs* reciprocal temperature in a sample of 4.2Ω cm *N*-type germanium. In many samples, the data were not sufficiently good to show region *A* as a region of distinctly steeper slope. Whenever such a steep slope could be distinguished, its activation energy was around 0.3 ev. The slope of region *B* differed widely from sample to sample, ranging from zero to the slope shown in Fig. 1.

In measuring the lifetime, it was often found that the slope of the decay curve shortly after injection* (short t) corresponded to a shorter lifetime than the slope far out on the tail of the decay curve (long t). High-level injection might be suspected, but this would require[2] that the small-t (high injection) value of τ_{plateau} be greater than the long-t (low injection) value of τ_{plateau}, which was contrary to most of the observed data. Furthermore, decreasing the size of the injection pulse had no effect, and calculated representative values of injection densities were small relative to majority carrier densities. A second possibility is that one region of the decay curve showed the effect of surface recombination, but this possibility seems unlikely in view of the insensitivity of either region of the curve to surface treatment. A third possibility is that the observed behavior resulted from traps, but the decay rate was not typical of traps, and shining light on the sample had no effect. Another interpretation is that the short-t value of τ is descriptive of the region very near the barrier, for at short t the carriers which are flowing across the barrier are those whose existence has been spent entirely in the region near the barrier. The long-t value would be descriptive, then, of the carriers and lifetime some distance away from the barrier. While this experiment does not purport to demonstrate conclusively the truth of this latter interpretation, it seems the most reasonable one and has been adopted in the analysis of the reverse-current data. Other interpretations could lead to slightly larger descrepancies between theory and experiment, but probably they would not be serious.

The significant conclusion to be drawn from these data is that the lifetime does reach a plateau at sufficiently low temperatures, as predicted by the Hall-Shockley-Read theory. We are inclined to interpret the slope of region *A* (see Fig. 1) as a measure of the depth of the recombination center in the forbidden band. In this connection, the effect of doping with iron or gold is of some interest, for in two such samples studied, this steep-slope region was considerably extended (see Figs. 2 and 3) and in both of these cases, as well as the undoped specimens, the slope was around 0.3 ev.

We believe that the various slopes of region *B* can be interpreted most easily by postulating the presence of a second recombination center about 0.1 ev deep, the various observed slopes then corresponding to different relative contributions by the two sets of recombination centers.[9] Since this belief is somewhat speculative, it has been relegated to Appendix I.

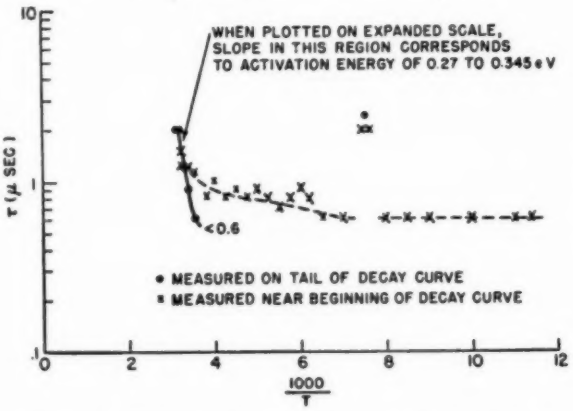

WHEN PLOTTED ON EXPANDED SCALE, SLOPE IN THIS REGION CORRESPONDS TO ACTIVATION ENERGY OF 0.27 TO 0.345 e V
<0.6
• MEASURED ON TAIL OF DECAY CURVE
× MEASURED NEAR BEGINNING OF DECAY CURVE
τ (μ SEC)
1000/T

FIG. 3. Log carrier lifetime *vs* $1000/T$, In dot junction on gold-doped germanium (14Ω cm *N*).

* This slope was nevertheless measured far enough out on the decay curve to be free from the effects of series resistance discussed by Lax and Neustadter, see reference 7.

[9] F. Rose (private communication) finds evidence for several additional levels. Since our data, which is perhaps more limited, does not require these additional levels for its explanation, we mention only the two and leave the presence or absence of other levels as an open question. The presence of other recombination levels is not unlikely in view of the many impurity levels found from resistivity and optical absorption measurements in the past few years.

Region C was not always so flat as in the sample described by Fig. 1. From the data accumulated to date, it seems safe to say that this plateau region has a temperature dependence which may be as low as zero but which does not exceed $\tau \sim T^{\frac{1}{2}}$. In higher resistivity samples, the beginning of plateau region C did not appear until liquid nitrogen temperature was approached (which is consistent with the interpretation given in Appendix I).

In the sample of Fig. 1, lifetimes measured near the beginning of the decay curve and thought to be characteristic of the region nearer the barrier were shorter than the lifetimes measured at longer t. In some samples, these lifetimes were nearly identical. In most samples the short-t lifetimes were somewhat smaller. In the iron and gold doped samples, the short-t lifetimes were slightly longer. These latter observations are pointed out only for completeness, and there is no desire to emphasize them until they are more completely understood.

Provided that the capture cross section of the recombination centers is different for electrons and holes and provided one knows the depth of the centers, it is theoretically possible to find whether the recombination centers lie closer to the valence or to the conduction band. One accomplishes this by noting the temperature at which the plateau is reached. In practice, the method is not very sensitive, and it has not yet proved possible to establish the band near which the centers lie.

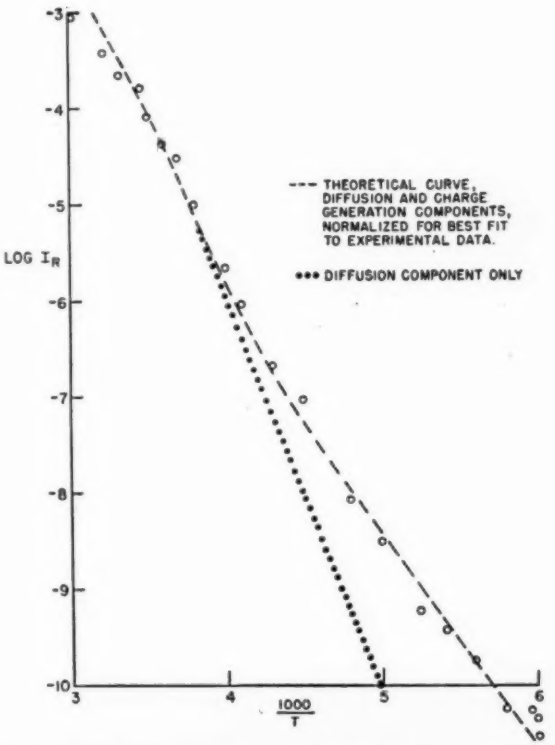


FIG. 4. Log reverse current (measured at low voltages and extrapolated to $V=KT/q$) vs $1000/T$, same sample as in Fig. 1.

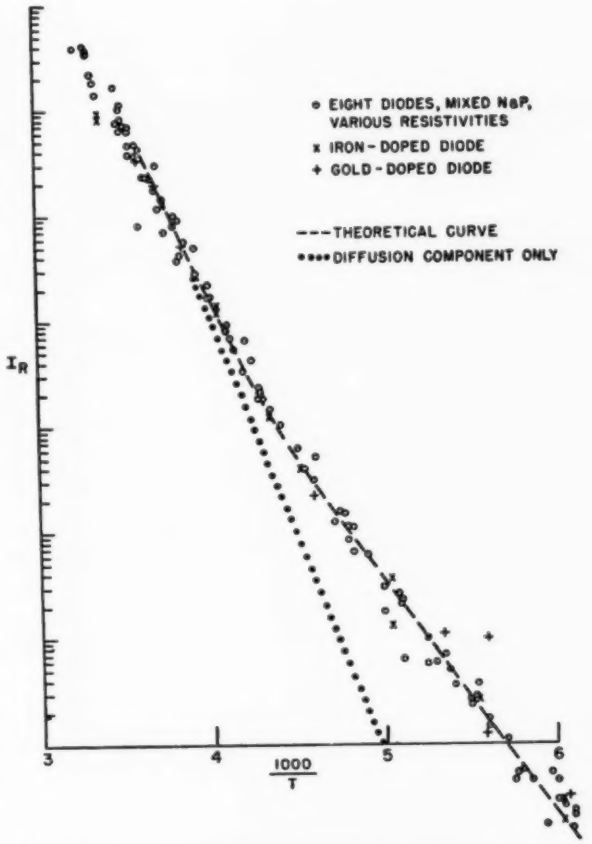


FIG. 5. Log reverse current vs $1000/T$ for ten samples, diffusion and charge generation regions normalized separately to show scatter of data.

Diode Reverse Current vs Reciprocal Temperature

Figure 4 shows reverse current vs reciprocal temperature for the same diode described by Fig. 1. In Fig. 5 the same data is plotted for all ten diodes, with the two regions of different slope normalized separately to show the extent of scatter of the data. The scatter is obviously too great to permit an accurate determination of the slope in the steep region, but the observed data in this region all lie within experimental error of a line whose slope is given by the usual diode diffusion theory.[4] This theory predicts that the slope should be given by the band gap energy (extrapolated to zero degrees Kelvin; we have used 0.75 ev though the fit would be equally good using the newer value of 0.785 ev[10]) modified by the $T^{\frac{1}{2}}$ dependence of N_v and N_c in the expression for the density of intrinsic carriers, $n_i^2=N_vN_c\exp(-\epsilon_G/kT)$, and further modified at the high-temperature end by the observed temperature dependence of minority carrier lifetime (see Appendix II). The temperature dependence of the diffusion constant can be neglected in the high-slope region within the accuracy at this experiment. The density of majority carriers remains constant in the temperature range of these measurements.

[10] F. J. Morin and J. P. Maita, Phys. Rev. **94**, 1525 (1954).

The significant feature of the shallow-slope region is that all diodes, within experimental error, exhibit the same slope, which corresponds to an activation energy of about 0.40 ev. While it cannot be denied that this could arise from some unknown reproducible surface effect, it is our feeling that this region is descriptive of charge generation in the space-charge region. Our reasons can be briefly stated: 1. The Hall-Shockley-Read recombination theory predicts such behavior. 2. The depth of the responsible generation centers as found from the slope of the reverse current data (about 0.32 ev if the band gap is taken to be about 0.72 ev and temperature dependences of these energies are neglected) is in good agreement with the depth of the deeper recombination centers as found from lifetime data (about 0.30 ev). 3. The magnitude of the current in this region is in excellent agreement with theory (see Table I). 4. The shape of the current-voltage characteristic is somewhat different in this region, and this different shape is similar to what one would expect if charge generation is responsible (see Appendix II).

Magnitude of the Reverse Current

For the ten diodes tested, the magnitude of the reverse current at $1000/T=3.5$ (region of steep slope) was compared with the diffusion model, and the magnitude of the reverse current at $1000/T=5.0$ (shallow-slope region) was compared with the charge generation model described in Appendix II. The results are given in Table I.[11]

It will be seen in Appendix II that the diffusion component is proportional to an expression $N_v N_c \exp(-\epsilon_G/kT)$ and the charge generation component is proportional to an expression $N_c \exp\{(-\epsilon_G+\epsilon_R)/kT\}$ or $N_v \exp\{(-\epsilon_G+\epsilon_R)/kT)\}$. These expressions collect the terms which cannot be measured by the present experiment, the exponential being included because of the unknown temperature dependence of the band gap. Now N_v and N_c are known from theoretical considerations to each vary as $T^{\frac{3}{2}}$, and the temperature dependence of the band gap can be described by $\epsilon_G=\epsilon_0-cT$. We may therefore use our data for the diffusion region to obtain a value for the expression $N_v N_c/T^3 \exp(c/k)$. This can be compared with the value for the same expression obtained in recent measurements of intrinsic resistivity in germanium[10] (equal to 3.1×10^{32}). In the case of the charge generation component, we should really compare our results with $N_v/T^{\frac{1}{2}} \exp(c/k)$ or $N_c/T^{\frac{1}{2}} \exp(c/k)$ (there is some question about the use of the exponentials here; their use depends upon to which band the recombination centers are most closely bound). We have compared them with $[N_v N_c/T^3 \exp(c/k)]^{\frac{1}{2}}$; the results will then give an experimental measure of the ratio of N_v to N_c.

The following information from Table I is of some interest, though it is to be qualified by whether or not one is willing to accept the proposed model:

1. For the diffusion component, the average of the ratios for the ten samples is 2.8 ± 1.3, which indicates fair agreement between this experiment and intrinsic resistivity data.

2. For the charge generation component, the average ratio is 1.3 ± 0.6, which indicates good quantitative agreement with the proposed model.

Comparison of Depth of Observed Center with Other Data

This experiment indicates the presence of a recombination/generation center about 0.3 ev from the valence or the conduction band in all samples tested.

W. W. Tyler has found iron levels from resistivity measurements at 0.34 ev above the valence band and 0.26 ev below the conduction band,[12] and the presence of these levels has been confirmed by R. Newman using infrared techniques.[13] Our results could be explained in terms of these same levels. They could also be explained by the copper level reportedly lying 0.3 ev above the valence band.[14] On the other hand, Dunlap has measured gold levels at 0.15 ev above the valence band and 0.2 ev below the conduction band[15] (values which are further supported by data of R. Newman[16]) which appears to contradict the similar behavior of the lifetime and reverse current in the two materials. There are three possibilities: 1. This particular gold sample accidentally contained iron. 2. Gold-doped germanium also contains a deeper center whose presence is masked in resistivity measurements by the shallower centers. 3. Both iron and gold bring with them another center, perhaps a lattice defect of some kind, which is the same in both materials and constitutes the recombination center.

CONCLUSIONS

1. The lifetime, if plotted *versus* temperature, reaches a plateau at low temperatures.

2. At intermediate temperatures, the slope of $\log\tau$

[11] A few of the methods adhered to in calculating the values in Table I may not be entirely obvious: 1. In calculating the width of the space-charge region corresponding to the applied potential of one millivolt, the capacitance extrapolated to the value of the electrostatic potential of the junction (estimated to be about 0.35 volt at $1000/T=5$) was used, because this was much larger than the applied potential. 2. The value of plateau lifetime used in calculating the charge generation component was that corresponding to the short-t lifetime (which is presumed to be more typical of the space-charge region) and that corresponding to the deep recombination centers only, since these are the effective charge generation centers. (From Appendix II, deeper centers correspond to a larger exponential factor in Eq. (12), resulting in a larger I_R from Eq. (11). In the samples used, this effect of the exponential factor far outweighed the opposing effect of the ratio of plateau lifetimes.) This plateau was estimated from the point at which region A (see Fig. 1) intersects region B (or the highest lifetime in region B if region A is too short to be distinguishable; the resulting ratio in Table I is indicated by the symbol $\geq$) using Fig. 6 as a guide. For the sample of Fig. 1, this would be about 10 μsec.

[12] W. W. Tyler and H. H. Woodbury, Phys. Rev. **96**, 874 (1954).
[13] R. Newman and W. W. Tyler, Phys. Rev. **96**, 882 (1954).
[14] J. F. Battey and R. M. Baum, Phys. Rev. **94**, 1393 (1954).
[15] W. C. Dunlap, Jr., Phys. Rev. **91**, 1282 (1953).
[16] R. Newman, Phys. Rev. **94**, 278 (1954).

versus reciprocal temperature varies from sample to sample, while at higher temperatures there is usually a steep region, having an activation energy of about 0.3 ev in these samples. One possible interpretation is that there are two levels of recombination centers present, one about 0.3 ev deep and the other very roughly 0.1 ev deep.

3. The slope and magnitude of log reverse current *versus* reciprocal temperature agrees with the usual diode diffusion theory at higher temperatures. At lower temperatures, there is a region of lower slope, reproducible from sample to sample. One possible interpretation is that this arises from charge generation in the space-charge region. If so, the generation centers in these samples are about 0.3 ev from the nearest allowed band. The magnitude of the current agrees with that calculated from this model.

ACKNOWLEDGMENTS

The author wishes to acknowledge the help and advice of many members of the laboratory staff, particularly R. N. Hall and W. W. Tyler. He is appreciative of the assistance of D. Locke. Samples used in this investigation were supplied by R. N. Hall, W. W. Tyler, W. C. Dunlap, Jr., W. E. Cady, and J. M. Mulhern, to whom the author expresses his indebtedness.

APPENDIX I

The Hall-Shockley-Read model for recombination of holes and electrons has been described elsewhere.[1,2] The results of this experiment can be explained by the presence of recombination centers of at least two energy levels, and it is therefore necessary to extend the theory to include more than one level. If the recombination centers act independently, the total recombination rate of two sets of centers at different levels will be simply the sum of the individual recombination rates, or

$$R_{\text{total}}=R_1+R_2. \tag{1}$$

Since $\tau=\partial n/R$,[1,2] where τ is the lifetime of a carrier, ∂n is the departure from equilibrium of the carrier concentration, and R is the recombination rate, it follows that

$$\frac{1}{\tau_{\text{two centers}}}=\frac{1}{\tau_1}+\frac{1}{\tau_2}, \tag{2}$$

where the subscripts refer to the two levels of recombination centers. The results of such a summation, for the extrinsic temperature range, are shown in Fig. 6, for different relative contributions (as described by the different plateau lifetimes τ_{10} and τ_{20}) by centers at 0.3 ev and 0.1 ev from either the valence or the conduction band.

It will be noticed that such relative contributions can produce a region at intermediate temperatures having a surprisingly straight slope, but a slope which will vary depending upon the relative densities of the

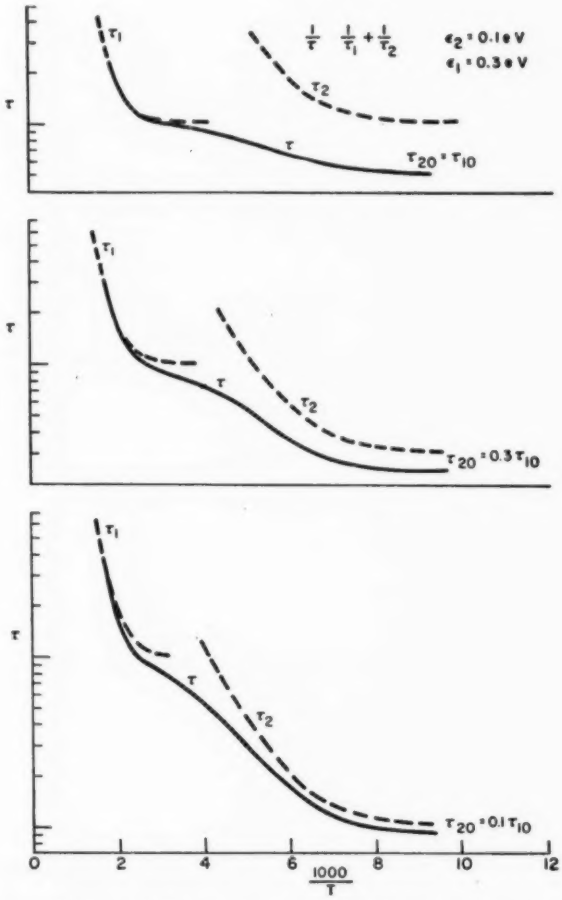


FIG. 6. Calculated effect on $\log\tau$ *vs* $1000/T$ if there are recombination centers at two levels, 0.1 and 0.3 ev deep.

two kinds of centers. This varying slope is characteristic of our experimental observations.

It will also be noted that the shallower centers, if observable, will cause the break to the plateau region to come at about the temperature observed in this experiment, and comparison with the Hall-Shockley-Read theory[1,2] will show that this break should occur at lower temperatures in higher resistivity samples, which was also observed.

It will be noted, furthermore, that if the density of the deeper centers is sufficiently increased, the effect of the shallower centers will no longer be observed. Such was the result in the iron- and gold-doped samples.

APPENDIX II

The reverse current of a junction diode is normally described by

$$I_R=eAn_{\min}\left(\frac{D_{\min}}{\tau}\right)^{\frac{1}{2}}, \tag{3}$$

where

$$n_{\min}=\frac{n_i^2}{n_{\text{maj}}}=\frac{N_vN_c\exp(-\epsilon_G/kT)}{n_{\text{maj}}}, \tag{4}$$

where the subscripts refer to minority and majority

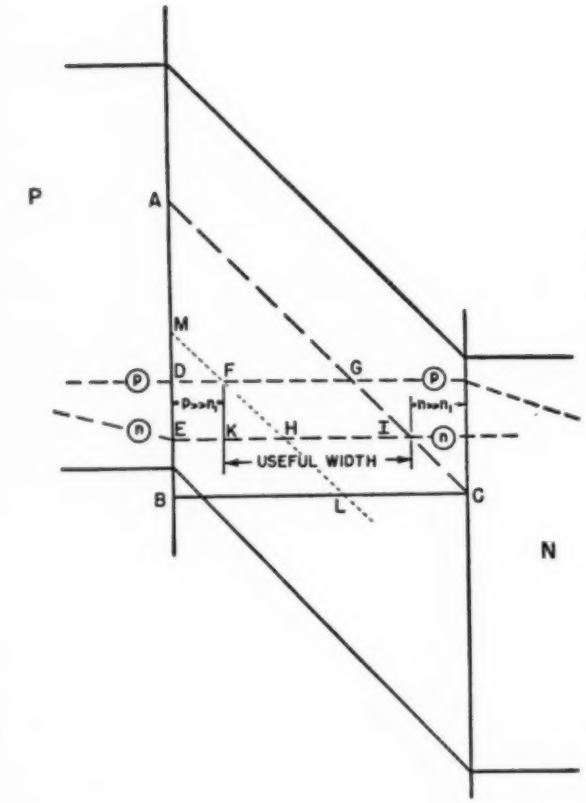


FIG. 7. Geometric construction for finding approximate width of useful region for charge generation.

carriers, e is the electronic charge, A is the area of the junction, n is the density of carriers, $D_{\min}$ is the diffusion constant for minority carriers, τ is the lifetime, n_i is the density of intrinsic carriers, N_v and N_c are the partition functions for the valence band and the conduction band, respectively, (each equal to $2(2\pi m^* kT/h^2)^{\frac{3}{2}}$ for a simple band, where m^* is the effective mass and h is Planck's constant), ϵ_G is the band gap, k is Boltzmann's constant, and T is the absolute temperature. There are really two components to this reverse current, corresponding to the two different minority carriers, holes and electrons, on the two sides of the junction, but one can be neglected relative to the other when the resistivities of the N and the P regions are substantially different. We shall refer to the component of reverse current defined by Eqs. (3) and (4) as the diffusion component.

There should also be a contribution arising from charge generation by the recombination centers in the space-charge region,[2] and the rate of this charge generation is given by

$$G=\frac{n_i^2-np}{\tau_N(n+n_1)+\tau_P(p+p_1)}, \tag{5}$$

where n and p are the densities of electrons and holes, respectively, in the space-charge region, τ_N and τ_P are the limiting values of lifetime† at very low temperatures if this region were N-type or P-type, respectively, and n_1 and p_1 are the densities of electrons and holes, respectively, if the Fermi level were at the energy level of the recombination centers. If a voltage V is applied to the junction, the quasi-Fermi levels in the space-charge region will be separated by this voltage V, so that

$$np=n_i^2\exp(-eV/kT). \tag{6}$$

We also know that $n_i^2=n_1p_1$, giving

$$G=\frac{n_1p_1\{1-\exp(-eV/kT)\}}{\tau_N(n+n_1)+\tau_P(p+p_1)}. \tag{7}$$

For V sufficiently large in the reverse direction, the field in the space-charge region will be sufficient to make $n\ll n_1$ and $p\ll p_1$. Generally, either $n_1\tau_N\ll p_1\tau_p$ or $n_1\tau_N\gg p_1\tau_P$, giving $G\cong n_1/\tau_P$ or $G\cong p_1/\tau_N$, respectively.

For V small, the analysis is more complicated. For example, if the recombination centers are near the conduction band, near one end of the space-charge region $n\gg n_1$ and near the other end $p\gg n_1$, while in the region between, $n\ll n_1$ and $p\ll p_1\ll n_1$ as before. This is made clear by Fig. 7, which is further described in the next paragraph. In the end regions of the space-charge region, the denominator of the expression for G will hence be much greater, and to a first approximation we can neglect the contribution to G of these end regions. By using straight line approximations for the energy band structure across the space-charge region, we can make a geometric construction which will give an approximate value for the fraction of the barrier width which contributes most of the generated charge.

The construction is given in Fig. 7. The P and N sides of the space-charge region are marked, as are the quasi-Fermi levels. The line AC gives the position of the recombination centers (nearer the conduction band than the valence band in this example). Line ML is drawn at a distance from the valence band equal to the spacing of AC from the conduction kand and is used for defining the region where $p\gg n_1$. The useful width is defined in accordance with the preceding paragraph and is appropriately labelled on the diagram. The useful fraction of the space-charge region for charge

† We use this nomenclature because the lifetime is a function of the region in the crystal and not of the type of carrier; in a given region, the lifetime of an excess electron is the same as that of an excess hole, provided injection is small. See reference 2.

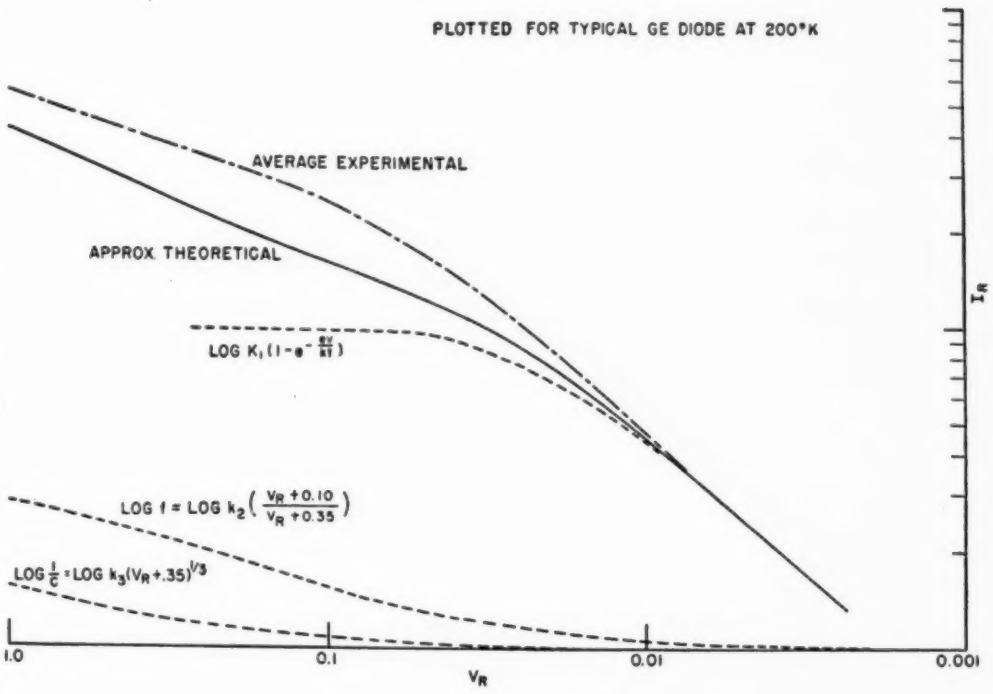


FIG. 8. Log reverse current *vs* log voltage when charge generation predominates over diffusion component, and comparison of approximate theory with experiment.

generation is then given by:

$$f\cong\frac{KI}{BC}=\frac{KH+HI}{BC}=\frac{KH}{BC}+\frac{LC}{BC}$$

$$=\frac{DE}{AB}+\frac{AM}{AB}=\frac{DE+AM}{AB}, \quad (8)$$

now

$$AB=V_R-V_{\Delta F}$$

$$AM=\epsilon_G-2\epsilon_R$$

and

$$DE=V_R.$$

Therefore,

$$f\cong\frac{V_R+\epsilon_G-2\epsilon_R}{V_R+V_{\Delta F}}.$$

A typical value‡ at 200°K in Ge would be $f\cong(V_R+0.10)/(V_R+0.35)$.

The reverse current arising from this generation will be given by

$$I_R=eGfWA, \quad (9)$$

where W is the width of the space-charge region. This width can be found by measuring the capacitance C, and it can be shown that, for germanium, $W(\text{cm})=1.42A/C(\mu\mu\text{f})$, independent of the impurity concentration gradient in the junction region. The width will, of course, be a function of the electrostatic potential across the space-charge region, which is equal to the applied reverse voltage plus $V_{\Delta F}$.

We have, then, for the component of I_R caused by generation in the space-charge region:

if $$n_1\tau_N\gg p_1\tau_p, \quad (10)$$

$$I_R=\frac{1.42eA^2}{C}f\left(\frac{p_1}{\tau_N}\right)\{1-\exp(-eV/kT)\},$$

and if $n_1\tau_N\ll p_1\tau_p$

$$I_R=\frac{1.42eA^2}{C}f\frac{n_1}{\tau_p}\{1-\exp(-eV/kT)\}, \quad (11)$$

where C is measured in $\mu\mu$f.

The resulting log-log plot of the $i-V$ characteristic as calculated for a typical Ge diode at 200°K is given in Fig. 8, which shows the contributions of the individual voltage dependent factors of Eq. (11), and which also includes an average of experimental data at this temperature.

The temperature dependence of this charge generation component is given by the temperature dependence of n_1 or p_1. (There will also be a slight temperature dependence caused by the variation of $V_{\Delta F}$ with temperature, but this can be neglected.) Now

$$n_1=N_c\exp\{(-\epsilon_G+\epsilon_R)/kT\}. \quad (12)$$

A similar expression gives p_1. These expressions correspond to a lower activation energy than the activation energy for the diffusion component of reverse current, and at a sufficiently low temperature one would therefore expect to observe a reverse current composed largely of charge generation in the space-charge region.

‡ $V_{\Delta F}$ can be measured experimentally by shining a bright light on the junction, with ohmic contacts placed at least a diffusion length from the edge of the illuminated region. If the ohmic contact is closer than this, the interpretation will depend upon the conditions at the contact. $V_{\Delta F}$ can also be estimated from theoretical considerations, knowing the resistivities on the two sides of the junction. In any case, an approximate value will suffice since I_R is not very sensitive to $V_{\Delta F}$.

Theoretical and Experimental Study of Electromagnetic Scattering by Two Identical Conducting Cylinders*

R. V. Row
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received December 7, 1953; revised July 6, 1954)

The scattering of a cylindrical TEM wave by two parallel, identical conducting circular cylinders is developed as a special case of a theoretical analysis which treats the scattering by an arbitrary array of cylinders. Only the case with the incident *E*-vector parallel to the axes of the cylinders is considered, and attention is focused on the mutual effects present when the cylinder diameter and spacing are comparable to a wavelength.

The approximations made in the theory are tested experimentally using 3 cm microwaves in a parallel plate region. Significant departures from the results of the independent scattering hypothesis as predicted by the theory have been confirmed experimentally.

1. INTRODUCTION

WITHIN recent years a number of papers[1–8] have been published dealing with various theoretical and experimental aspects of the diffraction of electromagnetic waves by a planar grating of parallel conducting wires. In the main, the results of the theoretical analyses have been directed towards obtaining information about gratings of elements whose size and spacing are either small compared with a wavelength, or spacings large enough to allow mutual effects between elements to be neglected. Several workers[4,6,7–9] have formulated their particular problems for quite general element size and spacings but have later specialized the results to the aforementioned cases. This paper reports a preliminary investigation into the effect of mutual coupling in a planar grating of parallel identical conducting cylinders where the cylinder diameters and spacing are comparable to a wavelength. In view of the interest that this problem has generated recently, it is interesting at this point to take a brief look at the work that has been done on it in the past.

It is interesting to note that about 60 years ago the development of reliable spark oscillators which radiated barely useful amounts of energy in the decimeter wavelength range stimulated interest in the problem of diffraction of electromagnetic waves by the planar grating. J. J. Thompson[10] and Rayleigh[11] both worked on limited aspects of the problem and H. Lamb[12] applied the technique of conjugate functions to the successful calculation of the transmission coefficient of gratings of closely spaced small cylinders or thin strips. His results were confirmed experimentally by Schaefer and Laugwitz.[13] Ignatowsky[9] was the first to develop a general theory of scattering from an infinite planar grating of identical elements using a formal solution of Maxwell's field equations satisfying the appropriate boundary conditions. The periodic nature of the boundary conditions allowed the field to be represented in a series of propagating and evanescent plane waves. Ignatowsky's work received little attention until quite recently when the mode or waveguide type of expansion has been used in connection with a variational principle to calculate the mode coefficients. Here again results have been presented only for small cylinders.

Concurrently with Ignatowsky, Zaviska[14] developed an analysis of diffraction from an arbitrary array of parallel cylinders by expanding the scattered field in a series of Hankel functions representing cylindrical waves radiating from each cylinder. The results of this analysis were applied to scattering by two small cylinders. However, the method of analysis is interesting and is essentially the method employed in this paper. It has an advantage in the directness with which various approximations may be introduced at the end of the rigorous analysis.

Wessel's theory is restricted to small wires from the outset, and as such, requires no futher mention here except to say that his results have been recently confirmed experimentally by Esau, Ahrens, and Kebbel.[2]

Twersky's[7] novel grating analysis by the use of "multiply scattered" waves is essentially an iteration process based on systematically improving the results of the independent scattering hypothesis. The method has not been applied in the case of large cylinders so far as this writer is aware. The problem discussed here is

* Research supported by Office of Naval Research contract.

[1] W. Wessel, Hochfrequenztechnik und Electroakustik **54**, No. 2, 62–69 (1939).

[2] Esau, Ahrens, and Kebbel, Hochfrequenztechnik und Electroakustik **54**, No. 2, 113–115 (1939).

[3] N. Marcuvitz, *Waveguide Handbook* (Massachusetts Institute of Technology Radiation Laboratory Series), Vol. 10, pp. 184.

[4] J. W. Miles, Quart. Appl. Math. **7**, 45 (1949).

[5] W. Franz, Z. angew. Phys. **1**, 9 416–423 (1949).

[6] J. Schmoys, Report EM-18, Math. Research Group, Washington Square College of Arts and Sciences, New York University (1951).

[7] V. Twersky. Report EM-34, Math. Research Group, Washington Square College of Arts and Sciences, New York University (1951); also, V. Twersky, J. Appl. Phys. **23**, 1099 (1952).

[8] W. E. Groves, Antenna Laboratory Report, Issue 179, Series 7, University of California (1952).

[9] W. von Ignatowsky, Ann. Physik **44**, 369 (1914).

[10] J. J. Thompson, *Recent Researches in Electricity and Magnetism* (Oxford University Press, New York, 1893), p. 425.

[11] Lord Rayleigh, Proc. Roy. Soc. (London), **79**, 399 (1907).

[12] H. Lamb, Proc. Roy. Soc. (London) **27**, 523–544 (1898).

[13] C. L. Schaefer and M. Laugwitz, Ann. Physik **21**, 587–594 (1906).

[14] F. Zaviska, Ann. Physik **40**, 1023 (1913).

first set up for scattering of an incident cylindrical TEM wave (with the electric vector oriented parallel to the cylinder axes) by an arbitrary configuration of parallel cylinders, and later specialized to the case of a planar grating of identical cylinders. Owing to the complexity of the results the problem is reduced a step further by considering only two identical cylinders. The choice of a cylindrical incident wave, rather than the conventional plane wave excitation, is to facilitate comparison of the theoretical results with measurements made on the diffraction of 3.2 cm microwaves by two cylinders in a parallel plate region.

2. OUTLINE OF THE THEORY

In the following only scalar scattering by circular cylinders is considered since this results in a mathematically tractable problem; although from Lamb's[12] work on the grating of closely spaced small wires or strips it is expected that for small scattering elements the precise form of their boundary is secondary in determining their scattered field in directions away from the source of radiation. The theory assumes a current distribution on the surface of each *perfectly* conducting cylinder; the total field is then calculated through the use of one of Green's theorems. Application of the boundary conditions gives a series of integral equations for the current on each cylinder which takes into account arbitrary excitation and coupling between all the elements. The unknown surface current on each cylinder is then expanded in a complex Fourier series whose coefficients may be evaluated using the usual orthogonality property of the trigonometric functions. The resulting system of linear algebraic equations in the unknown coefficients may be written as an infinite matrix equation. The problem then remaining is to solve this system of linear algebraic equations. Various methods of numerical solution may be used, depending on the number of terms and accuracy required in the final result. For small cylinders the terms off the principal diagonal are small, and the meaning of the term "small" may be evaluated readily in estimating the importance of higher mode currents contributing to the scattered field. The results obtained at this point in the analysis are similar in form to those of Zaviska[14] who started by assuming a spectrum of scattered cylindrical waves and determining the spectral amplitudes from a consideration of the boundary conditions. There is also a formal analogy to the results of Twersky's[7] multiple order scattering analysis.

This theory is readily specialized to the case of a plane wave incident on an infinite planar grating of small wires. If the effects of higher order current modes are neglected, this result becomes identical with that of Wessel[1] who considered a uniform current distribution on the surface of the wires in his analysis.

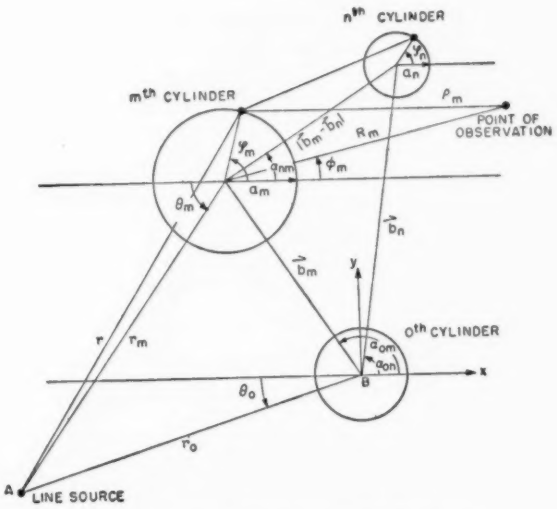


FIG. 1. Geometry for scattering from an arbitrary array of cylinders.

3. GENERAL THEORY (ARBITRARY CONFIGURATION OF PARALLEL CYLINDERS)

Figure 1 shows the general arrangement of line source and scattering cylinders. The axes of the cylinders and line source are all parallel to the z-axis so that all relevant electromagnetic field quantities may be derived from the single scalar quantity E_z, the electric field intensity in the z-direction, for convenience written as $\psi(x,y)$. With the customary time dependence $\exp(-i\omega t)$ suppressed throughout, and $k=2\pi/\lambda$, $\psi(x,y)$ satisfies

$$(\nabla_{x,y}^2+k^2)\psi(x,y)=0,$$

subject to the appropriate boundary conditions and the radiation condition at infinity.

If a Green's function $G(x,y;x',y')$ is defined as a solution of the inhomogeneous wave equation

$$(\nabla_{x,y}+k^2)G(x,y;x',y')=-\delta(x-x')\delta(y-y'),$$

and substituted into Green's scalar identity

$$\int(\psi\nabla^2G-G\nabla^2\psi)d\sigma'=\oint\left(\psi\frac{\partial G}{\partial n}-G\frac{\partial\psi}{\partial n}\right)dC',$$

the result

$$\psi(x,y)=\oint\left(G\frac{\partial\psi}{\partial n}-\psi\frac{\partial G}{\partial n}\right)dC',$$

is readily found, where the line integral is taken over a closed contour containing the source and all the cylinders. By imposing the boundary condition $\psi=0$ on the surface of each cylinder and making the convenient definition

$$\left.\frac{\partial\psi}{\partial n}\right|_{\text{on }n\text{th cylinder}}=\frac{1}{2\pi a_n}I_n(\phi_n),$$

where $I_n(\phi_n)$ may be considered as the surface current

on the nth cylinder, the previous result can be reduced to

$$\psi(\mathbf{r})=\psi^{\text{inc}}(\mathbf{r})+\frac{1}{2\pi}\sum_n\int_0^{2\pi} I_n(\phi_n)G(\mathbf{r},\mathbf{r}')d\phi_n, \quad (1)$$

where $\psi^{\text{inc}}(\mathbf{r})$ is the field that would exist at the point $\mathbf{r}$ if no scattering obstacles were present.

Application of the boundary condition $\psi(\mathbf{r})=0$ where $\mathbf{r}$ is on the surface of each cylinder and use of the appropriate two dimensional Green's function leads to the following integral equation for the surface current on each cylinder,

$$\psi^{\text{inc}}(\mathbf{r})\Big|_{\mathbf{r}\text{ on cylinder}} = -\frac{i}{8\pi}\sum_n\int_0^{2\pi} I_n(\phi_n)H_0^{(1)}(K|\mathbf{r}-\mathbf{r}'|)d\phi_n\Big|_{\mathbf{r},\mathbf{r}'\text{ on cylinders}} \quad (2)$$

The problem now is to find a set of $I_n(\phi_n)$ which satisfy this set of simultaneous integral equations. One method of solution is to expand the unknown function in a complete set of *ortho*-normal functions appropriate to the geometry of the particular problem and then to determine the resulting unknown coefficients. Following this method a natural choice here is to expand the surface current on each cylinder in the complex Fourier series

$$I_n(\phi_n)=\sum_{s=-\infty}^{\infty} a_{ns}\exp(is\phi_n).$$

By assuming the $I_n(\phi_n)$ to have a sufficiently regular behavior and using the orthogonal properties of the set of functions $\exp(is\phi_n)$, it is possible to reduce the problem of finding the a_{ns} to the solution of an infinite set of linear inhomogeneous simultaneous algebraic equations in which the a_{ns} are the unknowns.

If the mth integral equation is multiplied by $\exp(it\phi_m)$ (where t is any integer including zero), and both sides are integrated with respect to ϕ_m from 0 to 2π it follows that:

$$\gamma_{tm}=-\sum_n\sum_s K_{tmns}a_{ns} \quad (3)$$

where the definitions

$$\gamma_{tm}=\int_0^{2\pi}\psi^{\text{inc}}(\mathbf{r})\Big|_{\mathbf{r}\text{ on }m\text{th cylinder}}\exp(-it\phi_m)d\phi_m,$$

and

$$K_{tmns}=\frac{i}{8\pi}\int_0^{2\pi}\int_0^{2\pi}\exp(is\phi_n)H_0^{(1)}(k|\mathbf{r}-\mathbf{r}'|)\Big|_{\substack{\mathbf{r}\text{ on }m\text{th cylinder}\\ \mathbf{r}'\text{ on }n\text{th cylinder}}}\exp(-it\phi_m)d\phi_n d\phi_m \quad (4)$$

have been used.

The incident field characteristic of a uniform line source may be represented as

$$\psi^{\text{inc}}(\mathbf{r})=AH_0^{(1)}(k|\mathbf{r}|),$$

where A is a complex constant. With this choice of excitation the expressions for γ_{tm} and K_{tmns} as evaluated in the Appendix are

$$\gamma_{tm}=2\pi AJ_t(ka_m)H_t^{(1)}(k|\mathbf{r}_m|)\exp[-it(\theta_m+\pi)]$$

$$K_{tmns}=\frac{\pi i}{2}J_t(ka_m)\begin{cases} H_s^{(1)}(ka_m)\delta_{st}\Big|_{\mathbf{r},\mathbf{r}'\text{ on }m\text{th cylinder}} \\ J_s(ka_n)H_{t-s}^{(1)}(k|\mathbf{b}_m-\mathbf{b}_n|)\exp(it\alpha_{nm}+is\alpha_{nm})\Big|_{\mathbf{r},\mathbf{r}'\text{ on different cylinders}} \end{cases} \quad (5)$$

For any given values of the parameters k, a_n, θ_m and $|\mathbf{b}_m-\mathbf{b}_n|$; γ_{tm} and K_{tmns} may be evaluated using existing tables of the Bessel and Neumann functions.

Finally, the total field may be calculated from formula (1),

$$\psi^{\text{tot}}(x,y)=\psi^{\text{inc}}(x,y)+\frac{i}{4}\sum\sum a_{ns}J_s(ka_n)H_s^{(1)}(kR_n)\exp(is\phi_n). \quad (6)$$

Here the reader familiar with Twersky's[7] multiple scattering analysis will notice a formal analogy between Eq. (6) and his formula for the scattered field.

At this point it is convenient to show the connection between the analysis of this paper and the work of Wessel[1] on diffraction of a plane incident wave by a planar grating of small wires.

If the constant A is so chosen as to make the exciting field a plane wave incident from the direction θ_0 and the array is taken as a planar grating of identical cylinders then $b_m=m_b$,

$$\alpha_{0m}=\begin{cases}\alpha_0 & \text{for } m>0\\ \alpha_0\pm\pi & \text{for } m<0\end{cases}$$

and

$$a_m=a,$$

and from symmetry it is apparent that all the cylinders have the same current distribution except for the phase factor $\exp(ikmb\cos(\theta_0-\alpha_0))$ relative to the zeroth cylinder. Thus,

$$a_{ms}=\exp(ikmb\cos(\theta_0-\alpha_0))a'_{0s}$$

where a'_{0s} is the sth Fourier coefficient on the zeroth cylinder in the array. It is only necessary to calculate the current on the zeroth cylinder, thus for $m=0$ the system of Eq. (1) becomes

$$4i\exp\left(-it\left(\theta_0-\frac{\pi}{2}\right)\right)=H_t^{(1)}(ka)a_{0t}'+\exp(-it\alpha_0)\sum_{n\neq 0}\sum_s a_{0s}'\exp(iknb\cos(\theta_0-\alpha_0)+is\alpha_0)J_s(ka)H_{t-s}^{(1)}(k|n|b). \quad (7)$$

As t ranges through all positive and negative integers, an infinite number of linear simultaneous equations in the unknown a'_{0s} is generated. In principle this set of equations could be solved, but practically it is not feasible to solve for all the unknowns. For small ka and $b\gg a$, it may be seen from the behavior of the Bessel and Hankel functions involved that the dominant terms in the right-hand side of the previous expression are those with $t=s=0$. Thus Eq. (7) becomes

$$4i=\{H_0^{(1)}(ka)+\sum_{\substack{n=-\infty\\ \neq 0}}^{\infty}J_0(ka)H_0^{(1)}(k|n|b)\times\exp[inkb\cos(\theta_0-\alpha_0)]\}a_{00}'.$$

This is the same equation as obtained by Wessel for the current on the zeroth cylinder. Tables of the series $\sum_1^\infty J_0(nkb)$ and $\sum_1^\infty Y_0(nkb)$ corresponding to normal incidence have been computed by Ignatowsky[9] and Wessel.[1]

As far as coupling effects and their dependence on cylinder radius and spacing are concerned, a finite number of mode coefficients could be computed with a large amount of labor for the case of an infinite planar grating using Eq. (7). However, when it comes to comparing the theoretical results for the scattered field to experimental results, it is not feasible to use plane-wave excitation, and the system of equations for the mode coefficients for a line source excitation and the infinite grating are exceedingly complex. Hence, it is expedient to consider the simplest configuration for which mutual coupling effects may be calculated with a reasonable amount of labor. For these reasons the problem of scattering of a cylindrical wave by two identical cylinders has been chosen as an example to test the general theory.

4. TWO IDENTICAL CYLINDERS (NORMAL INCIDENCE)

Figure 2 shows the geometrical arrangement of source and cylinders for scattering of a cylindrical incident wave by two identical cylinders.

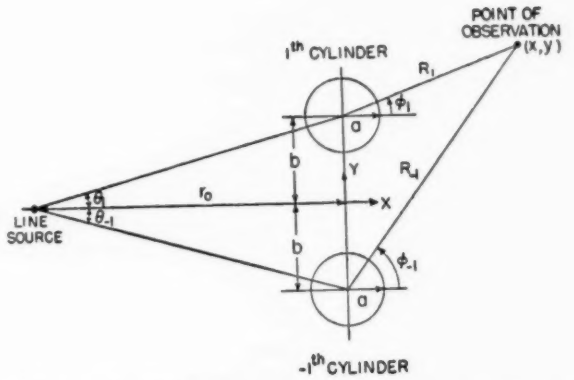


FIG. 2. Geometry for scattering from two identical cylinders (normal incidence).

The system of Eq. (3) for the unknowns $a_{\pm 1,s}$ (for $m=-1$) may be written

$$H^{(1)}(ka)a_{-1,t}+\exp\left(-\frac{it\pi}{2}\right)\{J_0(ka)H_t^{(1)}(2kb)a_{1,0}+\sum_{s=1}^{\infty}\exp(is\pi/2)J_s(ka)[H_{t+s}(2kb)a_{1,-s}+H_{t-s}^{(1)}(2kb)a_{1,s}]\}=4iH_t^{(1)}[k(r_0^2+b^2)^{\frac{1}{2}}]\times\exp[-it(\pi+\theta_{-1})]. \quad (8)$$

It is evident that

$$I_1(\phi_1)=I_{-1}(-\phi_1),$$

where $I_1(\phi_1)$ and $I_{-1}(-\phi_1)$ are the surface currents at mirror image points on the upper and lower cylinders. If the Fourier representation for the surface currents is combined with the above symmetry requirement, the following identity is readily established,

$$a_{1,s}=a_{-1,-s},$$

for all s.

The system of equations relating the unknowns obtained from Eq. (8) by using this statement of symmetry and allowing t to range through all positive and negative integers may be conveniently summarized in matrix form as,

$$K\cdot a=4i\lambda$$

where K is a square (infinite) matrix, whose elements are the coefficients of the $a_{\pm 1,s}$, and $4i\lambda$ is the infinite matrix of elements on the right-hand side of Eq. (8).

The rigorous solution of the problem requires the solution of an infinite matrix equation. Except for certain special cases (in this problem corresponding to no coupling, i.e., $kb\to\infty$), the solution to such an infinite matrix equation is not immediately obvious.

The best that can be done by way of a solution is to solve by numerical means, a finite number of the equations represented by Eq. (8). Putting aside any discussion of the rigorous justification of this procedure,

and granting that the approximation is good, how many, and which equations should be chosen for solution? Since the problem of scattering by two cylinders must reduce to that of scattering by a single cylinder in the limit of very large spacing between cylinders, the first question is most readily answered by using the simple well-known solution to the problem of the scattering of a plane wave by an isolated cylinder. The symmetry of the diagonal elements of the K matrix about the element corresponding to $t=s=0$ points to an obvious choice of equations as those with $t=0, \pm1, \pm2, \cdots, \pm n$, and unknowns with s ranging from $-n$ to $+n$. The scattered field (for the same orientation of the incident plane wave as in the two cylinder problem) from an isolated cylinder may be put in the form

$$\psi^{\text{scatt}}(r,\phi)=-\sum_{s=0}^{\infty}\epsilon_s i^{s+1}\sin\delta_s(ka)\times\exp[-i\delta_s)(ka)]H_s^{(1)}(kr)\cos s\phi$$

where

$$H_s^{(1)}(ka)=-iC_s(ka)\exp[i\delta_s(ka)]$$

and

$$\epsilon_s=\begin{cases}1, & s=0\\ 2, & s\neq 0\end{cases}.$$

In the far zone the amplitude of $H_s^{(1)}(kr)$ changes slowly with increasing s, and thus the change of amplitude of each term corresponding to a change in index s is essentially proportional to $|\sin\delta_s(ka)\cos s\phi|$. The term $\cos s\phi$ is one at most and from the tables for $\delta_s(ka)$ it is seen that $\sin\delta_s(ka)$ tends to zero with increasing s (for s greater than a certain integer). Thus, by reference to these tables, it is possible to pick out the greatest integer s for which any significant contribution will be made to the summation for the scattered field. This maximum integer may then be used as a guide in deciding upon the number of equations to be solved in the two cylinder problem. Thus, for $ka=3.0(2a/\lambda\sim1)$,

s	0	1	2	3	4	5	6	7
$\lvert\sin\delta_s(Ka)\rvert$	0.5680	0.7222	0.9496	0.4977	0.1426	0.02251	0.002094	0.000175

and $s=6$ is the largest index giving an appreciable term in the field summation. Hence in the two cylinder problem at least 6 modes must be solved for, corresponding to $t=0, \pm1, \pm2, \pm3, \cdots \pm6$. There is an enormous amount of labor involved in solving such a system of equations with complex coefficients, however, for a spacing of one wavelength between centers, such a system of equations has been solved exactly for the case $Ka=1.253$, and approximately for $Ka=2.0$ and 2.5, for t ranging between values determined by the above procedure.

Methods of solving such systems of equations on a desk calculator are well known,† but except for $ka<1.3$ the labor required to solve such a system by hand methods limits the usefulness of the theory. However, the solution of a "block" out of the matrix equation yields results in excellent agreement with the measurements to be discussed later.

In view of the computational difficulties encountered in solving a large number of linear algebraic equations with complex coefficients, it is desirable to have a simple approximation to the solutions for the unknowns a_1'$_s$ so that some of the major characteristics of the scattering by two cylinders may be more readily seen. The most obvious approximation, and the one to be discussed here, is suggested by the fact that the diagonal elements in the K matrix increase without bound with increasing t, and at the same time the off-diagonal terms tend to zero. Hence it seems reasonable as a first approximation to neglect all off-diagonal terms. The success of this approximation will be discussed later in connection with the experimental measurements.

The final expression for the scattered field using the diagonal approximation is readily found to be

$$\psi^{\text{scatt}}(R,\phi)=-\sum_{n=\pm1}\sum_{s=0}^{\infty}(-)^s\frac{J_s(ka)H_s^{(1)}[k(r_0^2+b^2)^{\frac{1}{2}}]H_s^{(1)}(kR_n)\exp[is(\phi_n-\theta_1)]}{H_s^{(1)}(ka)+J_s(ka)H_{2s}^{(1)}(2kb)}. \tag{9}$$

In the limit as kb and $kr_0\to\infty$ this expression (except for a constant) is identical with Seitz's[15] expression for the field scattered by an isolated conducting cylinder

The contribution of each mode in Eq. (9) depends principally on the quantity

$$\alpha_s(ka,kb)=\frac{J_s(ka)}{H_s^{(1)}(ka)+J_s(ka)H_{2s}^{(1)}(2kb)}. \tag{10}$$

As s increases, and for ka, $kb\gg1$ and ka, $kb<s$ the approximations for the Bessel and Hankel[16] functions yield for Eq. (11), the following asymptotic form,

$$\alpha_s(ka,kb)\simeq\frac{i}{\left(\frac{2t}{ka}\right)^{2t}+\left(\frac{1}{2\pi s}\right)^{\frac{1}{2}}\left(\frac{2s}{kb}\right)^{2s}},$$

† For the *exact* solution of 22 (real) linear equations Crout's method [P. D. Crout, Am. Inst. Elec. Engrs. **60**, (1941)] requires 5634 machine operations including a check column. For 1 percent accuracy ($ka=1.253$, $b/\lambda=1.0$) the Gauss-Seidel iteration method [See Whittaker and Robinson, *Calculus of Observations* (Blackie and Son Limited, London, England), fourth edition], requires 4 iterations totaling 2134 operations.

[15] W. Seitz, Ann. Physik **16**, 746 (1905), and Ann. Physik **19**, 544 (1908).

[16] G. N. Watson, *Bessel Functions* (Macmillan Company, New York, 1948), second edition, pp. 243, 198.

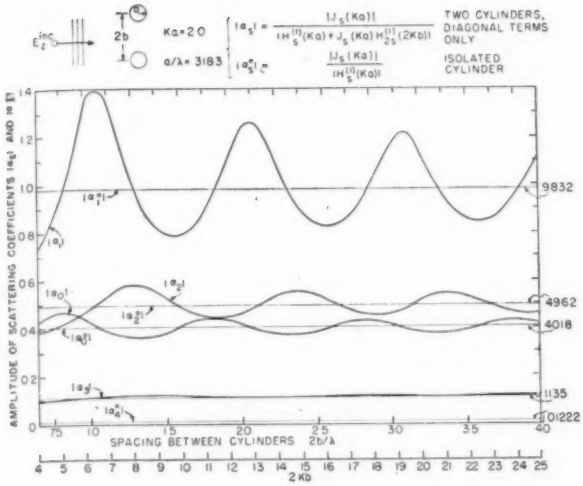


FIG. 3. Scattering coefficient $|\alpha_s|$ and $|\alpha_s^0|$ for two coupled cylinders (diagonal terms only), and an isolated cylinder as a function of spacing between centers for $Ka=2.0$ and normal incidence.

where the right-hand term in the denominator contains the coupling effect. From this last expression it is apparent that since $b \geqslant a$ that the two cylinders influence each other to a decreasing extent as the mode index s increases. This behavior is most strikingly demonstrated by the graphical plot in Fig. 3 of $|\alpha_s|$ computed from Eq. (10) as a function of spacing for $ka=2.0$. In addition $|\alpha_s^0|$, the corresponding quantity for an isolated cylinder is plotted for comparison along with $|\alpha_s|$. From this curve it is apparent that mutual effects tend to diminish slowly with increasing separation.

To compare the predictions of the diagonal approximation to measurable field quantities Eq. (9) has been used to calculate the scattered field. For a large separation between source and cylinders the following representation is used for the Hankel function $H_s^{(1)}(k(r_0^2+b^2)^{\frac{1}{2}})$

$$H_s^{(1)}(ku) \sim \left(\frac{2}{\pi ku}\right)^{\frac{1}{2}} \exp\left[i\left(ku-\frac{\pi s}{2}-\frac{\pi}{4}\right)\right] \times\left[S_s^{(1)}(ku)+0\left(\frac{1}{ku}\right)^p\right]$$

where $u=(r_0^2+b^2)^{\frac{1}{2}}$ and

$$S_s^{(1)}(ku)=\sum_{m=0}^{p-1} \frac{(-1)^m \Gamma(s+m+\tfrac{1}{2})}{(2iku)^m m!\Gamma(s-m+\tfrac{1}{2})}.$$

Since in practice $kr \gg s$ and $k=2\pi/\lambda$, the incident field at $x=x$ and $y=0$ is

$$\psi^{\text{inc}}(x,0) \simeq \exp(-i\pi/4)/\pi((r_0+x)/\lambda)^{\frac{1}{2}} \cdot \exp[ik(r_0+x)].$$

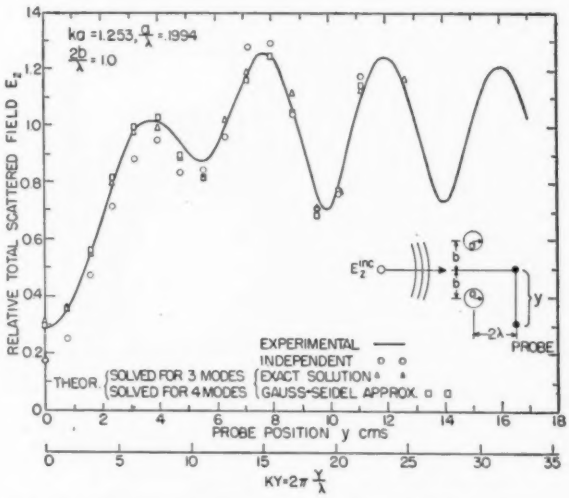


FIG. 5. Experimentally measured diffracted field from two cylinders and comparison with various theoretical approximations for $Ka=1.253$, as a function of probe position. (One wavelength between centers.)

Using the definition

$$\psi^{\text{tot}}=\psi^{\text{inc}}+\psi^{\text{scatt}},$$

and dividing this expression by the incident field at the reference point $x=x$, $y=0$ gives for the normalized total field.

$$\psi^{\text{tot}}{}_{\text{norm}}(x,y)=E_z=1-[(r_0+x)/(r_0^2+b^2)^{\frac{1}{2}}]^{\frac{1}{2}} \times\exp[ik(r_0^2+b^2)^{\frac{1}{2}}-ik(r_0+x)]\times \sum_{n=\pm 1} \sum_{s=-\infty}^{\infty} \times\exp(\pi is/2)\{\alpha_s(ka,kb)S_s^{(1)}[k(r_0^2+b^2)^{\frac{1}{2}}]H_s^{(1)}(kR_n) \times\exp[is(\phi_n-\theta_1)]\}.$$

This formula has been used in calculating the theoretical results labeled, diagonal terms only in Figs. 7, 9, 10, and 11.

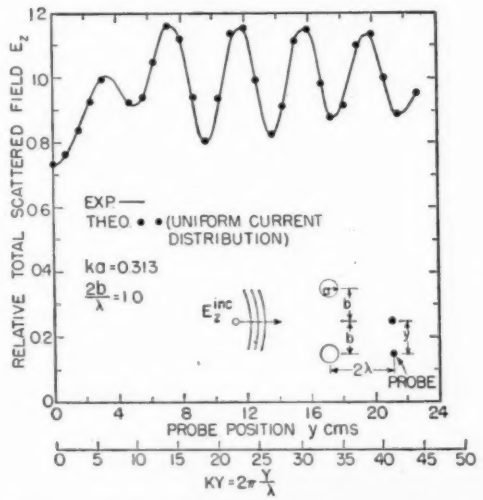


FIG. 4. Experimental and theoretical results of diffraction by two cylinders as a function of probe position for $ka=0.313$ and spaced 1.0 wavelength between centers.

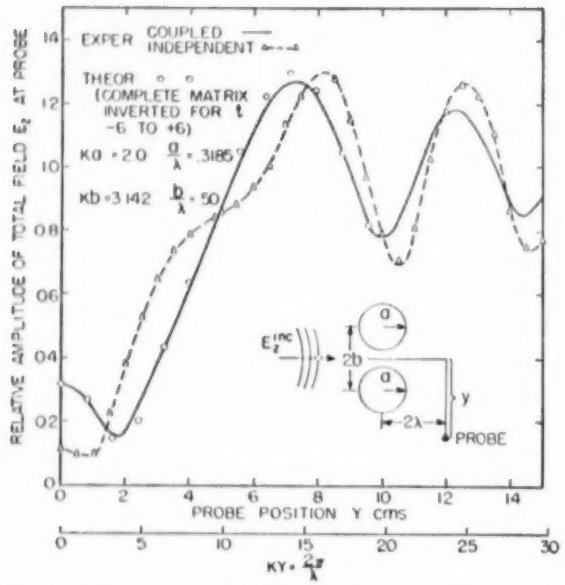

EXPER COUPLED
INDEPENDENT
THEOR
(COMPLETE MATRIX INVERTED FOR t -6 TO +6)
Ka = 2.0 a/λ = .3185
Kb = 3.142 b/λ = .50
RELATIVE AMPLITUDE OF TOTAL FIELD Ez AT PROBE
PROBE POSITION Y cms
KY = 2π/λ
PROBE

FIG. 6. Theoretical and experimental diffraction curves for scattering from two identical closely coupled cylinders, radius $a=0.3185\lambda$ and spacing between centers $=1.0\lambda$.

The factor $[(r+x)/(r_0^2+b^2)^{\frac{1}{2}}]^{\frac{1}{2}}$ represents an amplitude correction owing to the use of a line source in place of a plane wave, and the factor $ik(r_0^2+b^2)^{\frac{1}{2}}$ in the exponent plays the role of a phase correction factor. Inside the summation the term $S_s^{(1)}[k(r_0^2+b^2)^{\frac{1}{2}}]$ corrects for the line source excitation for each mode.

5. EXPERIMENTAL RESULTS

The parallel plate region and associated field-probing equipment described by the author in an earlier paper[17]

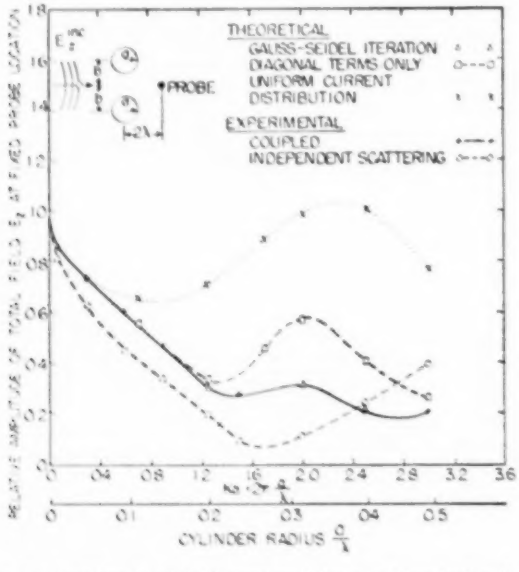


FIG. 7. Amplitude of field scattered by two identical cylinders, for fixed probe location and spacing $2b/\lambda=1.0$ as a function of radius of cylinders (line source).

[17] R. Row, J. Appl. Phys. 24, 12 (1953).

have been used to measure the electric field scattered by two identical highly conducting cylinders with a uniform line source for the primary excitation (at normal incidence) as sketched in Fig. 2.

These measurements are designed as a check on the validity of the approximations made in the theory developed in the preceding sections, and to compare the predictions of the independent scattering hypothesis with the field distributions actually measured. The results labeled "independent" scattering were determined in the following way. $E_z^{\text{inc}}(\mathbf{r})$ is the incident electric field at a point $\mathbf{r}$ and $E_{z_1}^{\text{tot}}(\mathbf{r})$ is the total electric field with cylinder 1 in place, then

$$E_{z_1}^{\text{scatt}}(\mathbf{r})=E_{z_1}^{\text{tot}}(\mathbf{r})-E_z^{\text{inc}}(\mathbf{r}).$$

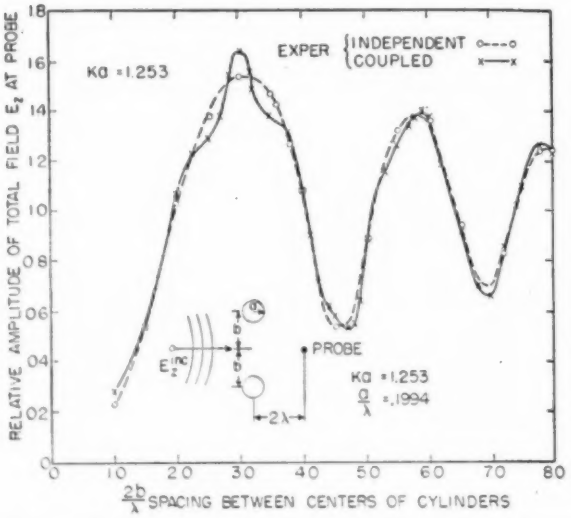


FIG. 8. Experimentally measured independent and coupled scattering from two identical cylinders with $Ka=1.253$ as a function of spacing. Probe position fixed.

Similarly, if cylinder 1 is removed and cylinder -1 is put in place:

$$E_{z_{-1}}^{\text{scatt}}(\mathbf{r})=E_{-1}^{\text{tot}}(\mathbf{r})-E_z^{\text{inc}}(\mathbf{r}).$$

According to the independent scattering hypothesis the total field with both cylinders in place is given by

$$\begin{aligned}E_z^{\text{tot}}(\mathbf{r})&=E_z^{\text{inc}}(\mathbf{r})+E_{z1}^{\text{scatt}}(\mathbf{r})+E_{z_{-1}}^{\text{scatt}}(\mathbf{r})\\&=E_{z_{-1}}^{\text{tot}}(\mathbf{r})+E_{z_{-1}}^{\text{tot}}(\mathbf{r})-E_z^{\text{inc}}(\mathbf{r}).\end{aligned}$$

The actual results presented required a measurement of amplitude and phase of the total scattered fields and the incident field. Brass cylinders $\frac{1}{8}$ inch thick and machined to the required diameters were used as the scatterers in the parallel plate region. In all cases the experimental results are reproducible to within 2 percent.

Figures 4 through 6 show the total field as measured by a probe moving along a line parallel to, and two wavelengths from (on the side away from the source), the line joining the centers of the cylinders; for a fixed center-to-center spacing of one wavelength and equal to 0.05λ, 0.20λ, and 0.318λ. Figure 7 shows the total

field at the point $x=2\lambda$, $y=0$ for a fixed center spacing of 1.0 wavelength as an increase from 0.0 to 0.477 wavelength. Figures 8 through 11 show the field at the same point as a function of spacing for a equal to 0.20λ, 0.24λ, 0.318λ, and 0.477λ.

In all cases the corresponding theoretical quantities are also shown on the same graph for comparison.

6. CONCLUSIONS

A study of Figures 5 through 11 shows that over a fairly large range or radii and spacings the independent scattering hypothesis may be used to predict large scale trends in the results. Thus for the probe fixed, and a constant spacing of 1.0 wavelength between

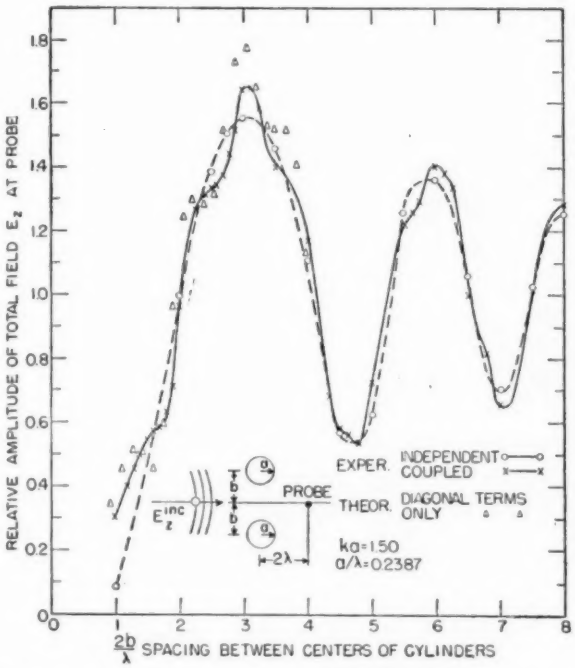


FIG. 9. Experimentally measured independent and coupled scattering from two identical cylinders with $ka=1.50$ as a function of spacing. Probe position fixed.

centers (see Fig. 7), the trend in the measured field as the radius increases is closely predicted by the independent scattering hypothesis for ka ranging from zero to approximately 1.5.

In all the measurements taken there is practically no detailed agreement between the independent scattering data and the results obtained with both cylinders present. As is to be expected the larger the radius, and the smaller the spacing, the poorer becomes the detailed agreement between the predictions of the simple independent scattering hypothesis and the corresponding measurements.

Attention will now be focused on the results of the theory developed in Sec. 4.

For small cylinders only the zeroth mode or uniform current mode is of significance in calculating the scattered field. The range of radii and spacing over which this mode alone is sufficient is determined primarily by the radius,‡ since here this parameter determines the number of modes required.

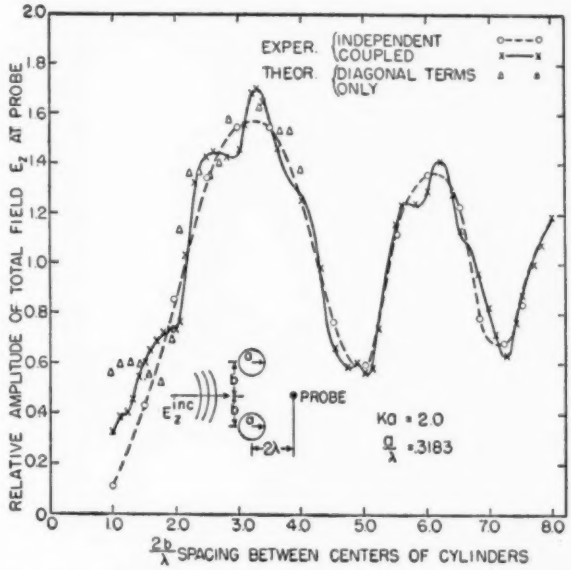


FIG. 10. Experimentally measured independent and coupled scattering from two identical cylinders with $Ka=2.0$, as a function of spacing. Probe position fixed.

Thus, with a fixed probe and a constant spacing of 1.0 wavelength between centers, Fig. 8 shows that the uniform current distribution gives excellent agreement with experiment for ka less than 0.3. The next step is

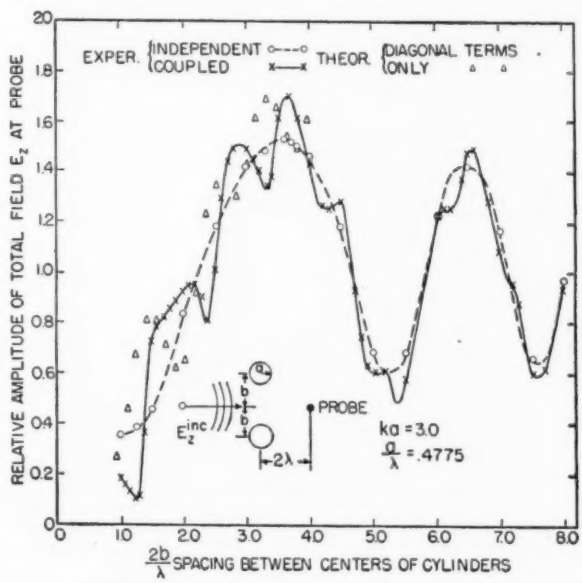


FIG. 11. Experimentally measured independent and coupled scattering from two identical cylinders with $Ka=3.0$ as a function of spacing. Probe position fixed.

‡ For cylinder diameters and spacings much less than a wavelength, the problem can be handled by the electrostatic approximation much as Lamb (see reference 12) treated the problem of a planar grating of small wires.

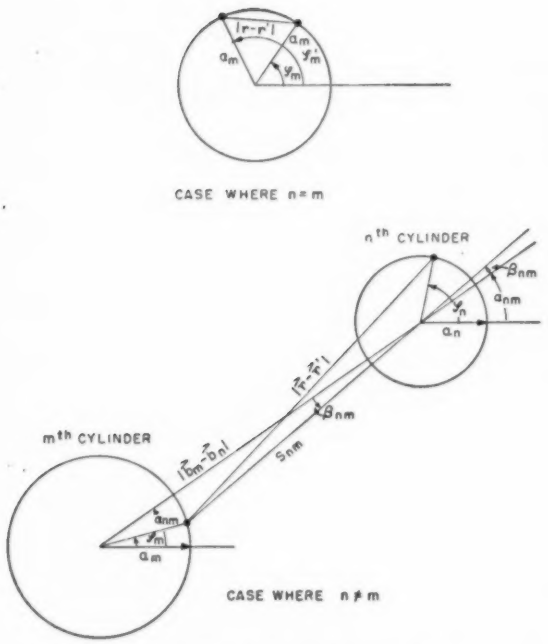


FIG. 12. Geometry for calculation of K_{tmns}.

to keep all the modes required for good convergence in the problem of diffraction by a single isolated cylinder. Owing to the amount of computing required, the effect of solving a finite number of equations for a fewer or greater number of modes has not been thoroughly investigated. However in one case, that for *ka* equal to 1.253 with a spacing of 1.0 wavelength (see Fig. 6), a finite "block" from the matrix equation was solved keeping at first three modes, and then four modes. Both cases lead to results in about equally close agreement with experiment. In all cases where the number of modes solved for was the same as required by the single cylinder problem, the detailed agreement between theory and experiment is excellent.

Without the use of large scale automatic computing machinery, it would be impractical to compute the solutions to the system of Eqs. (8) for any appreciable range of radii and spacings. Therefore, the diagonal approximation has been used to compute the total field at a point equidistant from each cylinder and two wavelengths behind the line joining their centers (see Figs. 7, 9 through 11).

From these results it is apparent that the diagonal approximation yields a satisfactory approximation to the detailed shape of the experimental curves. In comparing the experimental data with the theory for constant radius there is noted a progressive shift in the peaks and valleys of the theoretical curves towards smaller spacings as the radius increases, and for *ka* equal to 3.0 the initial valley in the experimental curve is absent in the theory. These effects indicate the increasing need for considering interactions between modes of different order in the theory as the radius is increased; that is, the mode coefficients must be obtained as the solutions to a block from the matrix equation.

Computations based on the diagonal approximation involve very little additional labor above that required by the independent scattering hypothesis and give much better results than this latter hypothesis, over a limited range, of course. Consequently, this method might be considered in examining other problems where the independent scattering hypothesis does not yield satisfactory results.

The author wishes to thank Professor J. E. Storer for his illuminating discussions on the theory, and Miss M. Tynan, Miss E. Miller, and Mrs. M. Amith for their assistance with the computations. Mr. E. Roffey modified the parallel plate apparatus to allow measurements to be made of the scattering by two cylinders.

APPENDIX

Evaluation of the Integrals γ_m and K_{tmns}

With reference to Fig. 1

$$r\Big|_{\text{on } m\text{th cylinder}} = [r_m^2 + a_m^2 - 2r_m a_m \cos(\theta_m + \pi - \phi_m)]^{\frac{1}{2}},$$

and

$$r_m = [r_0^2 + b_m^2 + 2r_0 b_m \cos(\theta_0 - \alpha_{0m})]^{\frac{1}{2}}.$$

Using the addition theorem for cylinder functions,

$$H_0^{(1)}(Kr) = \sum_{l=-\infty}^{\infty} J_l(Ka_m) H_l^{(1)}(Kr_m) \exp[il(\theta_m + \pi - \phi_m)];$$

then

$$\gamma_{tm} = A \sum_l \int_0^{2\pi} J_l(Ka_m) H_l^{(1)}(Kr_m) \exp(il(\theta_m + \pi) - il\phi_m - it\phi_m) d\phi_m,$$

where interchanging the order of summation and integration is assumed to be valid. The orthogonality relations for the trigonometric functions enable γ_{tm} to be evaluated explicitly, hence

$$\gamma_{tm}=2\pi AJ_t(Ka_m)H_t^{(1)}(Kr_m)\exp[-it(\theta_m+\pi)].$$

It should be noted that for cylinders with centers above the line $A-B$ in Fig. 1, θ_m is to be taken as positive, and for those with centers below $A-B$, it is negative.

The computation of K_{tmns} must proceed in two steps. The first case is that in which the indices m and n are equal, corresponding to r and r' lying on the surface of the same cylinder.

With reference to Fig. 12 for the case $n=m$,

$$|\mathbf{r}-\mathbf{r}'|=a_m[1-\cos(\phi_m-\phi_m')]^{\frac{1}{2}}.$$

Using the above addition theorem for the cylinder functions gives us

$$K_{tmns}=\frac{i}{8\pi}\sum_l\int_0^{2\pi}\int_0^{2\pi}\exp(is\phi_m-il\phi_m)J_l(Ka_m)H_l^{(1)}(Ka_m)\exp[il(\phi_m-\phi_m')]d\phi_m d\phi_m'=\frac{\pi i}{2}J_t(Ka_m)H_t^{(1)}(Ka_m)\delta_{st},$$

where

$$\delta_{st}=\begin{cases}1 & \text{for } s=t\\ 0 & \text{for } s\neq t.\end{cases}$$

We turn now to the evaluation of K_{tmns} when $n\neq m$. From Fig. 12 it may be seen that

$$|\mathbf{r}-\mathbf{r}'|=(s_{nm}^2+\alpha_n^2-2s_{nm}a_n\cos[\pi+\alpha_{nm}-\phi_n+\beta_{nm}])^{\frac{1}{2}}.$$

Again, applying the same addition theorem,

$$K_{tmns}=\frac{i}{8\pi}\sum_l\int_0^{2\pi}\int_0^{2\pi}\exp(is\phi_n-il\phi_m)J_l(ka_m)H_l^{(1)}(ks_{nm})\exp[il(\pi+\alpha_{nm}-\phi_n-\beta_{n\beta})]d\phi_n d\phi_m.$$

Carrying out the ϕ_n integration, we get

$$K_{tmns}=\frac{i}{4}\int_0^{2\pi}\exp(-il\phi_m)J_s(ka_n)H_s^{(1)}(ks_{nm})\exp[is(\pi+\alpha_{nm}+\beta_{nm})]d\phi_m.$$

Both s_{nm} and β_{nm} are dependent on ϕ_m so that in order to perform the remaining integration, the more general addition theorem

$$H_s^{(1)}(Ks_{nm})\exp(is\beta_{nm})=\sum_{q=-\infty}^{\infty}J_q(Ka_m)H_{s+q}^{(1)}(K|\mathbf{b}_m-\mathbf{b}_n|)\exp[iq(\alpha_{nm}-\phi_m)]$$

must be used whence it may readily be shown that

$$K_{tmns}=\frac{\pi i}{2}J_t(ka_m)J_s(ka_n)H_{t-s}^{(1)}(k|\mathbf{b}_m-\mathbf{b}_n|)\exp(-it\alpha_{nm}+is\alpha_{nm}).$$

Summarizing,

$$\gamma_{tm}=2\pi AJ_t(ka_m)H_t^{(1)}(kr_m)\exp[-it(\theta_m+\pi)]$$

$$K_{tmns}=\frac{\pi i}{2}J_t(ka_m)\begin{cases}H_s^{(1)}(ka_m)\delta_{st}, & n=m\\ J_s(ka_n)H_{t-s}^{(1)}(k|\mathbf{b}_m-\mathbf{b}_n|)\exp(-it\alpha_{nm}+is\alpha_{nm}), & n\neq m.\end{cases}$$

Trajectory-Tracer Study of Helix- and Band-Type Postdeflection Acceleration*

GEORGE C. SPONSLER†
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts
(Received September 25, 1954)

A high-speed automatic electron-trajectory tracer was employed to study the relative loss in deflection sensitivity and linearity caused by continuous resistive helical and conventional band-type cathode-ray tube postdeflection accelerators. A total of 144 trajectories was traced for six different electrode configurations immersed in a wedge-type electrolytic tank that simulated a particular tube envelope. Beam-deflection angles of 5.5°, 13.5°, 21.5°, and 30° were employed with maximum potential differences of 0, 2.3, 4.8, 7.3, 9.8, and 12.3-kv postdeflection acceleration and a 2.7-kv initial electron energy. The relative error of the resultant trajectories plotted is estimated at one percent. Of the configurations tested, the performance of the continuous resistive helix was far superior in deflection sensitivity and slightly better in deflection linearity. The results also indicate that a linear, stepped voltage distribution $[(V_0/N)\cdot n]$, although poorer than the continuous-helix distribution, may give better performance than that resulting from a stepped, geometrical voltage distribution (V_0, $V_0/2$, $V_0/4$, etc.).

I. INTRODUCTION

POSTDEFLECTION acceleration, also termed "after-acceleration,"[1] is employed in cathode-ray tubes to permit beam deflection at low field intensities together with subsequent high final electron energies at the phosphor screen. This double objective is obtained by applying some form of accelerating electrode or electrodes on the cathode-ray tube envelope between the deflection region and the viewing screen. These accelerating electrodes are operated at high electrostatic potential differences with respect to the electron beam energy as measured in the deflection region. This method of acceleration permits the use of low-power beam-deflection amplifiers (for either electrostatic or magnetic deflection), yet provides high light-spot brightness on the viewing screen which results from the high final electron velocity. Unfortunately, the accelerator electrodes necessarily behave like an electrostatic electron-optical lens which acts to bend the deflected beam back toward the central axis of the tube. Expressed another way, a lens action is produced which results in a loss of deflection sensitivity and also in a loss in deflection linearity. (Other lens aberrations are introduced as well, but these are second-order effects.)

A number of different postdeflection electrode configurations have been suggested.[2] The conventional electrode system employs conductive bands painted upon the tube envelope with potential steps applied to button connectors inserted through the tube wall.[3] A more unusual proposal was made by Schwartz[4] who suggested that the accelerating electrode be painted upon the glass envelope in the form of a continuous resistive helix across the two ends of which would be applied the full potential difference to create the accelerating field. Such an electrode, it was thought, should have the theoretical advantage that it should behave as a weak thick lens and should cause less loss in deflection sensitivity than the conventional band-type accelerators. Unfortunately, the electrostatic potential distribution within the cathode-ray tube envelope resulting from such a continuous resistive helix can only be calculated for comparatively simple geometries (e.g., for a cone closed by a portion of a sphere[5]), and the resultant electron trajectories require tedious numerical computations. Therefore, it was decided to simulate various electrode configurations in

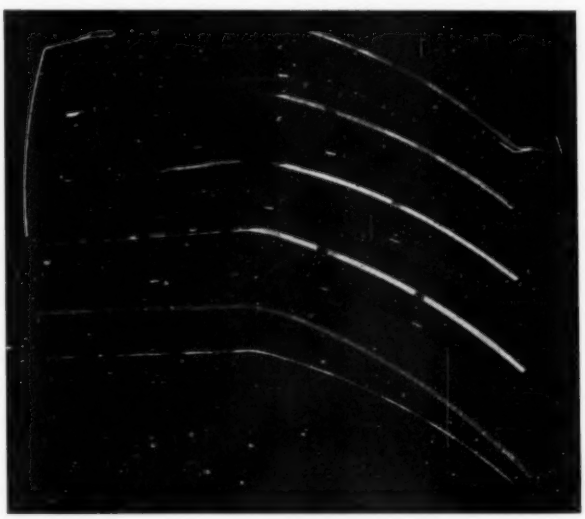


FIG. 1. 1-, 2-, 3-, 4-, and 88-band gold-plated electrodes and holders.

* The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

† Staff Member, Lincoln Laboratory, Massachusetts Institute of Technology.

[1] J. R. Pierce, *Theory and Design of Electron Beams* (D. Van Nostrand Company, Inc., New York, 1949), p. 137.

[2] E. Schwartz, Ferseh, A. G., 1, 1938.

[3] T. Soller, *et al.*, *Cathode Ray Tube Displays*, M.I.T. Radiation Laboratory Series (McGraw-Hill Book Company, Inc., New York, 1948), Vol. 22, p. 65.

[4] E. Schwartz, Germany, Patent No. 2 123 636.

[5] G. C. Sponsler, "Electrostatic potential distribution for the helically wound postdeflection accelerator of a conical cathode-ray tube," Technical Report No. 58, Lincoln Laboratory, Massachusetts Institute of Technology (January 20, 1954).

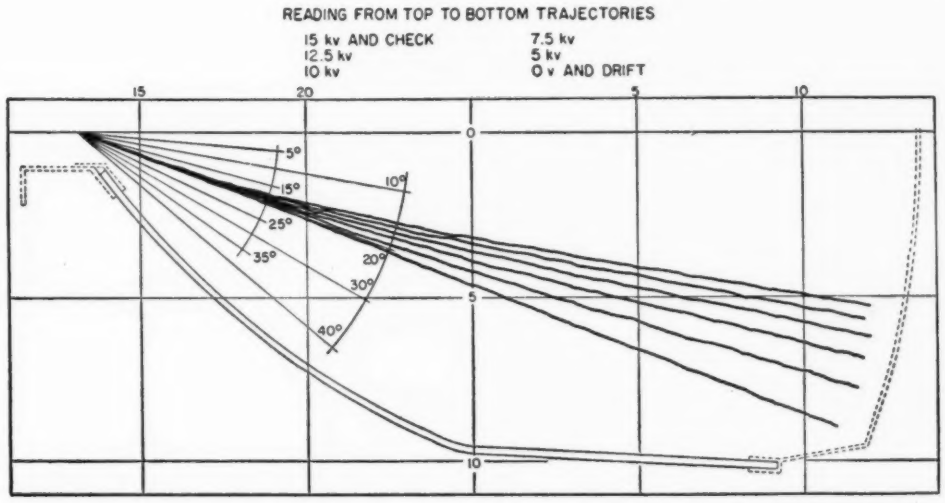


FIG. 2. Typical 21.5° deflection angle cathode-ray trajectories for different total postdeflection potential differences.

a wedge-type electrolytic tank[6] and to employ an automatic electron-trajectory tracer to trace the electron trajectories for a variety of imposed conditions.

A complete description of the automatic trajectory tracer is to be published elsewhere.[7,8] However, it should be mentioned that the device is similar to that of Sanders[9] in that it employs an electromechanical differential analyzer[10] to integrate the equations of motion of the electrons.

$$\frac{d^2x}{dt^2} = -\frac{e}{m}E_x(x,y); \qquad \frac{d^2y}{dt^2} = -\frac{e}{m}E_y(x,y). \tag{1}$$

The tracer differs from the English one principally in that it employs a novel type of dual-axis servo "function generator" to position the pickup probe (which represents the instantaneous electron position in the simulated field of the electrolytic tank and which measures the x- and y-field components at that point) and operates much faster (1 in./sec as compared to 0.4 in./min). Using this apparatus, it was quite simple, once the tracer had been "checked-out," to trace 250 separate trajectories in less than two days' total running time.

II. SIMULATOR PROCEDURE

The envelope outlines of a particular set of cathode-ray tubes (see Figs. 1 and 2) scaled to three-fifths size, were simulated in the electrolytic tank by gold-plated metal and gold-sputtered formed Pyrex glass electrodes. Banded accelerators of 1, 2, 3, 4, and 88 bands were simulated, as well as the continuous resistive helix of 88 turns. The electrolytic wedge tank simulated the electrostatic field of the cathode-ray tube which possessed rotational symmetry about its central axis. Gold was employed to combat polarization of the electrolyte.[11] For the same reasons, fine platinum wires were employed for the pickup probe. Mechanical misalignment of the probe wires and differing channel gains could be compensated by an ingenious cross-coupling network in the computer which, in essence, performed an electrical linear transformation of the projected field components. For further details as to the actual tracer design, reference should be made to the literature cited.[7,8] Two photographs of the apparatus are presented in Figs. 3 and 4.

Employing a concentric-cylinder, central-force field auxiliary electrolytic tank as a calibration aid,[12] the initial electron (i.e., probe) velocity was scaled to represent 2700 electron volts. Initial deflection angles of 5.5°, 13.5°, 21.5°, and 30° were set into the computer. The actual voltages applied to the tank were scaled to represent maximum postdeflection acceleration potential differences of 0, 2.3, 4.8, 7.3, 9.8, and 12.3 kv, which corresponded to 2.7, 5, 7.5, 10, 12.5, and 15 kv taken with reference to the cathode-ray tube cathode. These maximum accelerating voltages were applied across a resistor voltage divider by means of a variable autotransformer (Variac). For the single-band electrode, the full voltage was applied between the single accelerator and the neck electrode which represented the deflection region of the cathode-ray tube. For the multiple 2-, 3-, and 4-band electrodes, taps were taken on the voltage divider so that the voltages applied to the electrodes between the viewing screen and the

[6] M. Bowman-Manifold and F. H. Nicoll, Nature **140**, 39 (1938).
[7] Logan, Welti, and Sponsler, "Analogue study of electron trajectories," (to be published).
[8] G. C. Sponsler, "An automatic electron-trajectory tracer study of helix- and band-type postdeflection acceleration," Technical Report No. 64, Lincoln Laboratory, Massachusetts Institute of Technology (May 26, 1954).
[9] K. I. Sanders, *et al.*, Proc. Inst. Elec. Engrs. (London) **3**, 169 (1952).
[10] This differential analyzer was developed at the Dynamic Analysis and Control Laboratory, Massachusetts Institute of Technology, and was adapted to the present use by George Welti, and Benjamin Logan, of that laboratory, and the author.
[11] P. A. Einstein, Brit. J. Appl. Phys. **2**, 49 (1951).
[12] See references 7 and 8 for operational details.

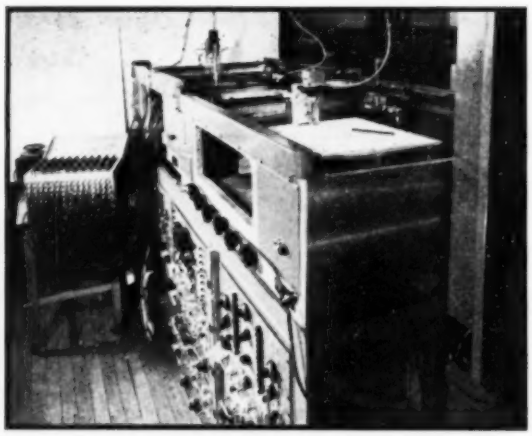


FIG. 3. Over-all view of trajectory tracer.

neck of the cathode-ray tube were stepped down by successive factors of $\frac{1}{2}$ (e.g., for 4 electrodes, the applied voltages were V_0, $V_0/2$, $V_0/4$, and $V_0/8$). Unfortunately, the lack of scheduled running time prevented other voltage ratios from being employed.

The 88-band electrode and the 88-turn continuous-helix electrode were simulated by the same gold-sputtered 88-band formed Pyrex glass strip (see Fig. 1). When operated as a banded electrode, each individual gold strip was connected to a corresponding point on a linear voltage divider resistor bank in such a manner that the voltage of the nth strip was $(V_0/88) \cdot n$, where the electrodes were numbered starting at the neck of the cathode-ray tube. The particular voltage corresponding to each strip (which would be present if the electrode structure were an actual continuous resistive helix) was calculated numerically for what would be the intersection points of the actual helix with the 88 circular bands. The proper voltage distribution was then obtained from taps on a resistor voltage divider composed of 88 precision resistors especially wound to provide the desired voltage drops. The resistors were wound to an accuracy of one-half of one percent.

III. RESULTS AND OBSERVATIONS

Using the automatic electron-trajectory tracer, the individual trajectories were traced at a speed of about 1 in./sec. This high rate reduced the errors arising from the integration drift and amplifier drift in the equipment but was not high enough to introduce dynamic errors in the various servomechanical components. To further reduce the relative error, one particular solution (the single-electrode 30° curve) was employed as a check solution before running each trajectory set for the other electrode configurations. The first trajectory run in each of these sets was employed as a subsidiary check solution for that particular set of trajectories.

A discussion of equipment performance error is included in the other papers.[7,8] Here, it should be noted that the individual curve displayed a reproducibility of better than one-half of one percent. The principal relative errors arose from drift of the cathode followers which amplified the probe voltages and from the effect of electrolyte evaporation which lowered the water level at a rate of 0.005 in./min. It is estimated that the over-all relative error of the trajectories is one percent, but the absolute error is unknown. The actual trajectories traced are similar to that reproduced in Fig. 2 which represents a 21.5° deflection-angle solution. The trajectories do not extend up to the tube envelope, since the probe had to be halted in time to avert collision.

To analyze the curves, the intercepts of the individual trajectories with the abscissa marked "10" on the tracer diagram were tabulated. All the trajectories had become linear before reaching this "10" abscissa line as by then they had been accelerated through the greater portion of the applied postdeflection field. Hence, using the "10" abscissa intercepts rather than the actual intercepts at the screen of the cathode-ray tube is not thought to affect the relative behavior or the conclusions drawn from the trajectories. In the case of the 30° deflection-angle curve, one or two trajectories on each diagram had to be extended by dashed straight lines in order to obtain an intercept with the "10" abscissa. This extension is permissible, since each curve had already attained its linear characteristics before the initial extension point was reached. The graphs in Figs. 5 to 7 plot the actual deflection intercept d *versus* the tangent of the deflection angle. The family parameter is the maximum applied accelerating voltage.

Study of the curves shows that the continuous-helix characteristics is everywhere above all the other curves on each graph. (Unfortunately, the data for the continuous helix at the 5° deflection angle are missing. This omission was caused by an accident when the probe was bent before the very last set of runs.) From

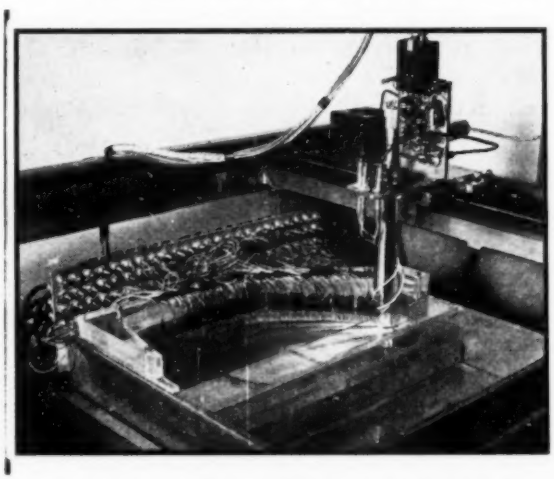


FIG. 4. Electrolytic tank with probe and 88-band electrode in position.

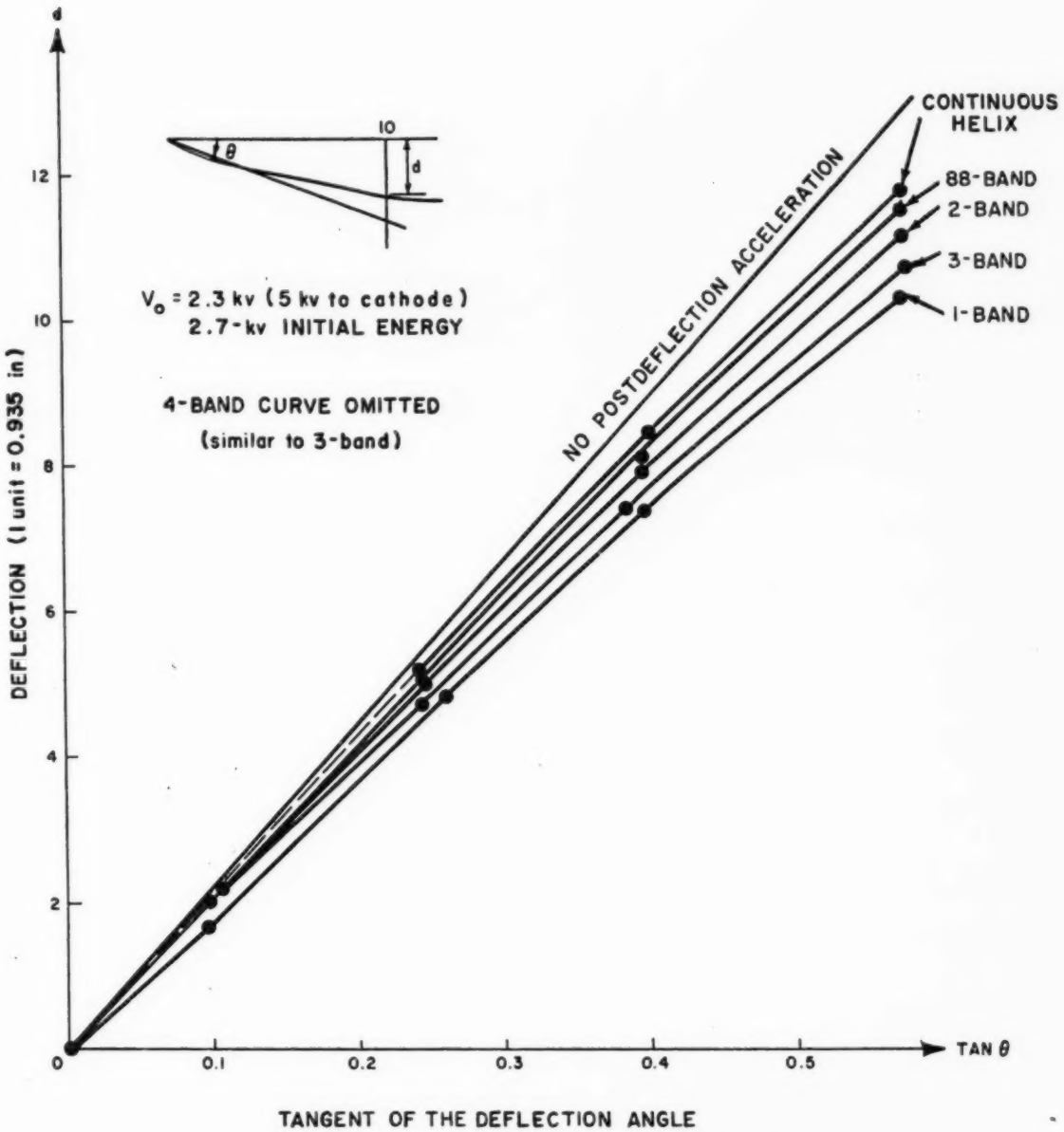


FIG. 5. Deflection *vs* deflection angle tangent, 2.3-kv acceleration.

which observation it may be stated that the deflection sensitivity is better for the continuous resistive helix than for any and all the other configurations.

The order of performance of the various configurations is somewhat interesting. In order of decreasing performance, they would be listed as continuous helix, 88-band electrode, 2-electrode, 3-electrode, 4-electrode, and (least desirable) single-electrode band accelerators. The position of the 3- and 4-electrode results between the numbers 1 and 2 is an interesting point. In a sense, the actual voltage steps chosen for the various electrodes of the 3- and 4-band configurations were incorrectly selected if it had been desired that their performance be intermediate between that of the 2- and the 88-band configurations; i.e., if it were desired to approach the performance of the 88-band configuration as a (sort of) limit for band-type electrodes. For such performance, the voltages applied to the 3- and 4-band electrodes ought not to have been stepped down "geometrically" by a factor of $\frac{1}{2}$ as they were, but rather voltages applied given by a formula similar to the 88-band electrodes' linear or "proportional" voltage distribution, i.e.,

$$V_n = \frac{V_0}{N} \cdot n, \tag{2}$$

where V_n is the voltage applied to the nth electrode, V_0 is the maximum potential difference applied, and N is the total number of electrodes. (Thus, for the

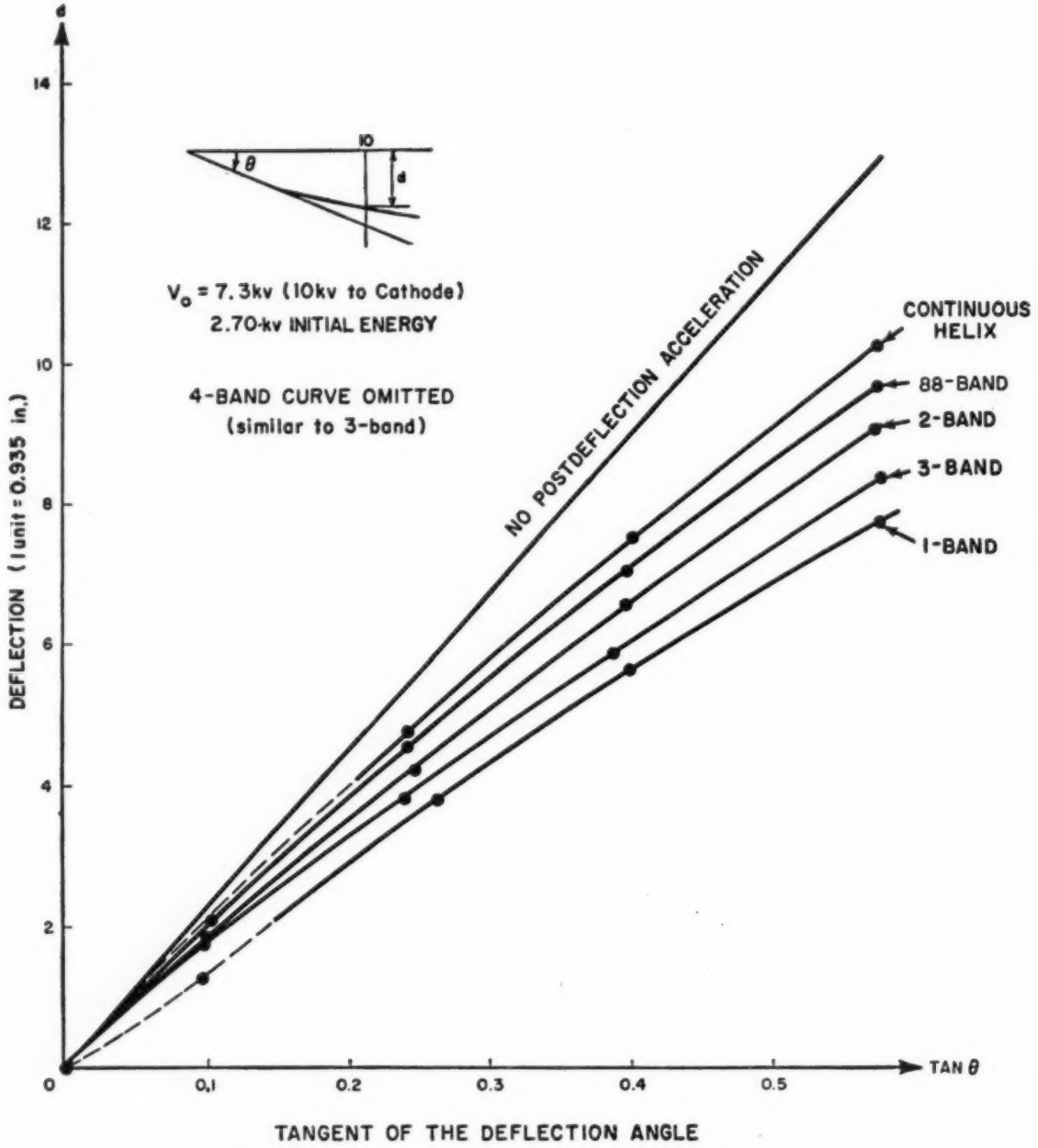


FIG. 6. Deflection *vs* deflection angle tangent, 7.3-kv acceleration.

4-electrode structure, the applied voltages would have been $V_0, 3V_0/4, V_0/2$ and $V_0/4$.) The poorer performance of the "geometrically" stepped voltages may indicate that the linear or proportional-type distribution, of which the 1-, 2-, and 88-band configurations are members, is better than the so-called "geometrical" one employing the factor $\frac{1}{2}$, of which the 3- and 4-band types are representative. Much more data would be required to confirm or deny this supposition.

Comparison of the relative deflection *linearity* of the various curves is slightly arbitrary. If one accepts the loss of deflection sensitivity and then asks what the deflection linearity is with this particular sensitivity accepted, none of the curves is really too bad. This argument is aided if the best straight-line approximation is fitted to each of the individual curves as could also be done to demonstrate the loss of deflection sensitivity. Then the comparison of the relative deflection linearity might be made by comparing the various, say, mean-square deviations away from these straight-line approximations. By this criterion, there is little to be gained in the choice of any of the various electrode configurations as regards deflection linearity.

If, however, deflection linearity is taken to mean the *curvature* of the various characteristics away from the zero potential straight-line reference (which is plotted

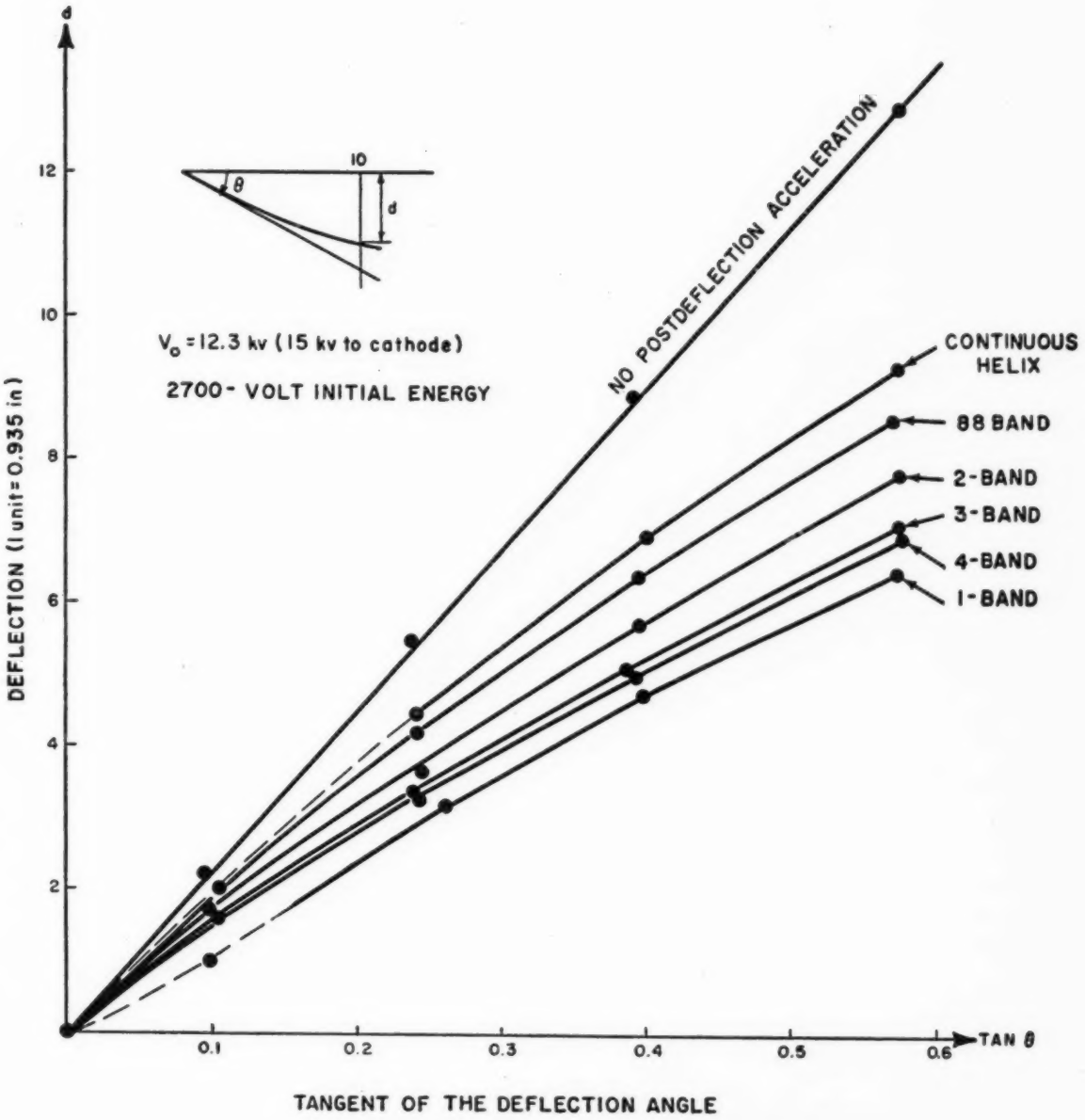


FIG. 7. Deflection *vs* deflection angle tangent, 12.3-kv acceleration.

on all graphs to aid comparison), then obviously the continuous resistive helix possesses the least curvature, and hence displays the best deflection linearity. The behavior of the other configurations follows in the same order as before.

Two observations ought to be made with reference to these graphs. First, there is actually little difference between the characteristics of the 3- and 4-electrode accelerators. This is of interest in relation to the argument concerning the proportional or linear *versus* the geometrical-type voltage distributions. In both the 3- and 4-band electrodes, half the total applied potentials apply between the last two bands in each case; hence, similar performance might be expected. Second, the 5° set of the single-electrode graphs are in error. This error was demonstrated by the customary drift check of the tracer which was made after each set of runs and, in this particular case, indicated substantial drift amounting to perhaps 10 percent. Similar drifts were encountered in no other cases.

Finally, it might be noted that the use of nonlinear deflection amplifiers possessing slightly rising gain characteristics could be employed to compensate partially for the loss in both deflection sensitivity and linearity resulting from the use of any of the types of postdeflection acceleration tested.

IV. CONCLUSIONS

It would, of course, have been desirable to study a greater variety of both electrode configurations and

other potential distributions. However, on the basis of the study made, it can be stated categorically that the continuous resistive helix gave the finest performance with respect to both deflection sensitivity and deflection linearity.

It would also seem to be indicated, on the basis of the relatively limited data taken, that the linear or proportional voltage distribution $[(V_0/N)\cdot n]$ gives better performance than the stepped or geometrical voltage distribution (V_0, $V_0/2$, $V_0/4$, etc.) in which fully one-half the accelerating field is applied between the last two electrodes next to the viewing screens.

Use of the continuous helix eliminates the extra bulb connectors needed for multiple-band electrode tubes and hence simplifies the cathode-ray tube glassworking requirements and construction. Practical continuous resistive helices have been made in the laboratory[13] and have performed creditably in actual cathode-ray tubes. At present, it is believed that the continuous resistive helix represents one of the best methods of postdeflection acceleration.

[13] At Tektronics, Inc., Portland, Oregon, and at Massachusetts Institute of Technology.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 6 JUNE, 1955

Performance of the Nonmagnetic Radio-Frequency Mass Spectrometer Tube*

T. C. Wherry and F. W. Karasek
Phillips Petroleum Company, Research and Development Department, Bartlesville, Oklahoma

(Received August 30, 1954; revised manuscript received January 17, 1955)

The nonmagnetic radio-frequency principle of mass selection appears attractive because it promises to result in a mass spectrometer of simple construction and rugged nature. Accurate evaluation of the use of this principle is difficult because of the lack of published performance data. In an effort to obtain performance data a 5-stage Bennett-type tube with stage separation of 5–9–4–7 cycles has been constructed and operational variables studied theoretically and experimentally.

The experimental results are encouraging. They indicate the tube is capable of sufficient stability, resolving power, and sensitivity to be used for many analytical and research problems. Electronic circuits required to achieve a reasonable stability are relatively simple.

I. INTRODUCTION

SEVERAL types of nonmagnetic radio-frequency mass spectrometer tubes have appeared in the literature. Among the most recent are those of Redhead[1,2] and Bennett.[3] Of the two, more work appears to have been done with the tube proposed by Bennett. Three stage tubes have been used by several people for various purposes[4–6] with varying degrees of success. Quite recently Townsend has described its use in an instrument designed for upper atmospheric research.[7,8] The data presented here concern a five stage tube designed according to this principle of mass selection proposed by Bennett. The stage separation is 5–9–4–7 cycles.

II. THEORETICAL CONSIDERATIONS

The principle may be explained briefly by reference to Fig. 1 which shows a schematic diagram of a 5-stage tube. The flow of electrons from the heated filament into the collector cage "3" is maintained constant by a control grid. Ions are formed inside "3" by electron bombardment. These ions, drawn out and collimated by grids 4 and 5, are accelerated into the mass selector section by grid 6. An rf stage is formed by grids 6, 7, and 8 with the rf potential on the center grid. Because of the accelerating potential, one particular ionic mass will traverse this stage so as to acquire maximum energy from the rf voltage. Others will acquire lesser amounts or even lose energy. A constant field drift space following this stage is so dimensioned that an integral number

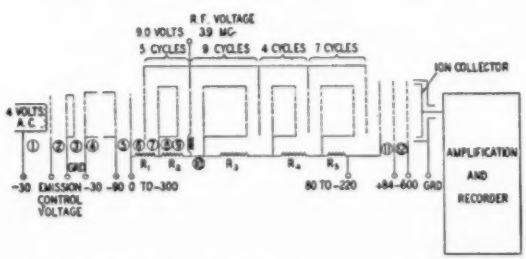


FIG. 1. Schematic of 5-stage nonmagnetic rf mass spectrometer tube.

* Presented in part before the Second Annual Meeting of ASTM E-14 Committee on Mass Spectroscopy, New Orleans, Louisiana, 1954.

[1] P. A. Redhead, Can. J. Phys. **30**, 1–13 (1952).
[2] P. A. Redhead, J. Appl. Phys. **24**, 331 (1953).
[3] W. H. Bennett, J. Appl. Phys. **21**, 143–149 (1950).
[4] R. F. Potter, "An analysis of background currents in bennett rf mass spectrometer." Thesis submitted to graduate school of University of Maryland, 1951.
[5] V. Keilin, Progress Report 24–3 Jet Propulsion Laboratory, California Institute of Technology.
[6] M. K. Testerman, AD-16212 Armed Services Technical Information Agency, June, 1953.
[7] J. W. Townsend, Rev. Sci. Instr. **23**, 538–541 (1952).
[8] J. W. Townsend, Elec. Eng. **25**, 470 (1953).

of rf cycles will occur during the time the resonant ion travels from grid 7 to grid 10. This enables the resonant ion to gain maximum energy from the second rf stage. All ions arrive at grid 11 with the resonant ion distinguished by having the greatest energy. A retarding potential on this grid will allow only the ion with the highest energy through to the collector. Grid 12 suppresses secondary electrons. The ion current is amplified and recorded by conventional means.

Since the rf voltage tends to increase the velocity of the resonant ion, R_1, R_2, R_3, R_4, and R_5 provide step-back voltages to counteract the rf voltage. By maintaining the velocity of the resonant ion essentially constant, an ion will resonate according to

$$M=0.26Va/s^2f^2,$$

where Va=accelerating voltage, s=spacing in centimeters between grids in a stage, f=frequency in megacycles, and M=mass of ion.

This provides a very simple, qualitative picture of the mass selection principle. To understand the complete operation of the tube a more rigorous quantitative analysis is necessary. No general equation with simple parameters can be devised to describe the motion and energy gained for any particular ion through the system. However, it is possible to derive a general equation containing a few variables and by a series of lengthy calculations obtain the data necessary to plot curves which reveal the quantitative functioning of a particular tube under set conditions.

Derivation of a usable general equation begins with the equation giving the energy gained by an ion as it passes through a single stage in the tube:

$$\Delta W_{RF}=Ve\frac{v}{s}\left[\int_{t_1}^{t_2}\sin(wt+\theta)dt-\int_{t_2}^{t_3}\sin(wt+\theta)dt\right] \quad (1)$$

where V=rf voltage (peak), e=charge on ion in electron volts, v=velocity of ion through stage (assumed constant), s=grid spacing, w=angular frequency of rf, θ=entrance phase angle of rf as ion enters first stage, t_1, t_2, t_3=times at which ion passes first dc, rf, and last dc grids, respectively.

For the first stage, one may take $t_1=0$ as the time the ion passes the first dc grid. Then: $t_1=0$, $t_2=s/v_1$, $t_3=2s/v_1$, where v_1 is the velocity of the ion due to accelerating voltage.

For the second stage, t_1 is the time to pass the first stage and the first drift space, or:

$$t_1=\frac{2s}{v_1}+\frac{d_1}{v_2},\quad t_2=\frac{2s}{v_1}+\frac{d_1}{v_2}+\frac{s}{v_2},\quad t_3=\frac{2s}{v_1}+\frac{d_1}{v_2}+\frac{2s}{v_2}.$$

For the nth stage,

$$t_1=\frac{2s}{v_1}+\frac{d_1}{v_2}+\frac{2s}{v_2}+\frac{d_2}{v_3}+\frac{2s}{v_3}+\cdots+\frac{2s}{v_{n-1}}+\frac{d_{n-1}}{v_n},$$

$$t_2=t_1+\frac{s}{v_n},\quad t_3=t_2+\frac{s}{v_n}=t_1+\frac{2s}{v_n}.$$

Equation (1) may be integrated to yield the following equation for the gain in energy produced by the rf voltage:

$$\begin{aligned}\Delta W_{RF}=Ve\Bigg[&\frac{v_1}{sw}\left\{\cos(0+\theta)-2\cos\left(\frac{sw}{v_1}+\theta\right)\right.\\&\left.+\cos\left(\frac{2sw}{v_1}+\theta\right)\right\}+\frac{v_2}{sw}\left\{\cos\left(\frac{2sw}{v_1}+\theta+\frac{d_1w}{v_2}\right)\right.\\&-2\cos\left(\frac{2sw}{v_1}+\theta+\frac{d_1w}{v_2}+\frac{sw}{v_2}\right)\\&\left.+\cos\left(\frac{2sw}{v_1}+\theta+\frac{d_1w}{v_2}+\frac{2sw}{v_2}\right)\right\}+\cdots\\&+\frac{v_n}{sw}\left\{\cos\left(\frac{2sw}{v_1}+\theta+\frac{d_1w}{v_2}+\frac{2sw}{v_2}\right.\right.\\&\left.+\frac{d_2w}{v_3}+\frac{2sw}{v_3}+\cdots+\frac{d_{n-1}w}{v_n}\right)\\&-2\cos\left(\frac{2sw}{v_1}+\theta+\cdots+\frac{d_{n-1}w}{v_n}+\frac{sw}{v_n}\right)\\&\left.\left.+\cos\left(\frac{2sw}{v_1}+\cdots+\frac{d_{n-1}w}{v_n}+\frac{2sw}{v_n}\right)\right\}\Bigg].\end{aligned} \quad (2)$$

In using this equation, all the variables must be fixed. One then computes the energy gain through the first stage assuming a constant ionic velocity. This approximation holds quite well because a step-back voltage is used. Since the step-back voltage is intended to maintain the resonant ion velocity constant, its

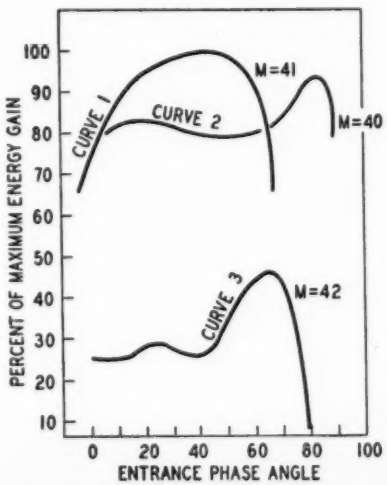


FIG. 2. Energy gain of ions through mass selector-mass 41 resonant.

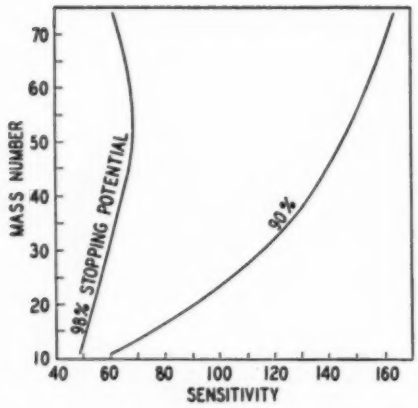


FIG. 3. Calculated values of mass number *vs* ion sensitivity.

value is calculated to provide an energy loss just equal to the energy gained from the rf voltage by the resonant ion entering at the optimum entrance angle of 46.43 degrees. To continue one must then know v_2 which is calculated from the difference between the energy gained from the rf voltage and that lost from the step-back voltage. This process is carried out until the ion has traversed the five stages. Another calculation can then be made by changing only the variable under study and repeating the same steps. All the data in Figs. 2, 3, and 4 were obtained in this manner.

Figure 2 shows the rf phase angle when an ion enters the first stage plotted against the percent of maximum energy gain an ion receives after traversing the entire tube. Curve 1 illustrates the energy gain of ions of the resonant mass, 41, for various entrance angles. It reveals that the energy gain falls off rapidly as resonant ions enter at other than the optimum angle of 46.43 degrees. It also reveals the "duty cycle" of the tube. If the stopping potential were set so ions gaining from 96 to 100 percent the maximum energy were allowed through to the collector, it would be possible for 37/360, or about 10 percent, of the available ions of mass 41 to reach the collector.

Curve 2 of Fig. 2 shows the energy gained by the mass 40 ion when the tube was adjusted to resonate

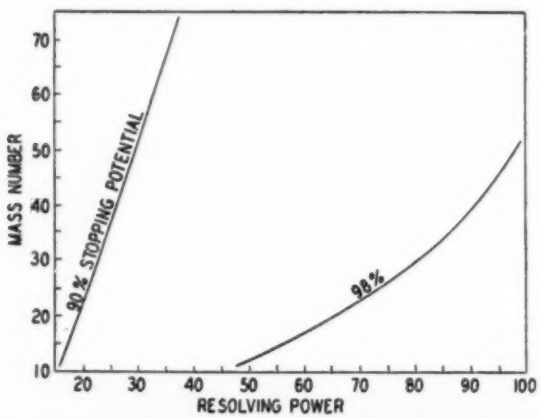


FIG. 4. Calculated values of mass number *vs* resolving power.

mass 41 ions. It indicates one must adjust the stopping potential at least above 94 percent to resolve mass 41 from mass 40.

Curve 3 of Fig. 2 indicates that the energy gained by the next heavier mass is considerably lower than the lighter mass.

These three curves do not show the energy gain of "harmonic" ions. These are ions whose velocity under a given accelerating voltage is such that they traverse the entire five stages in some very nearly integral number of cycles other than 5–9–4–7. Experiment has shown such "harmonic" ions do not gain more than 85 percent of maximum energy. Since the tube is normally operated at a stopping potential of 94 percent or greater, they constitute no problem.

Figure 3 reveals the relationship between arbitrary sensitivity units and mass at different stopping po-

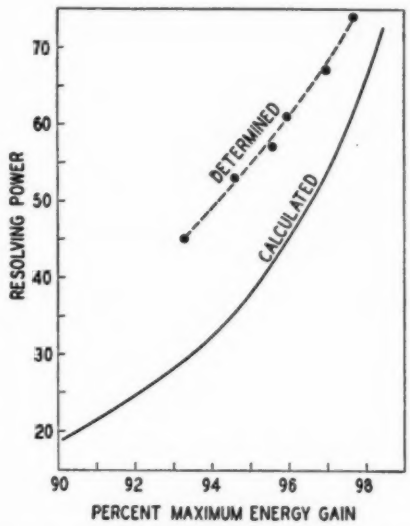


FIG. 5. Calculated *versus* determined resolving power values at mass 20 ion.

tentials. It shows a discrimination against the lighter masses.

Figure 4 reveals the relationship between resolving power and mass at different stopping potentials. Again the discrimination against the lighter masses may be noted. It should be pointed out that the resolving power definition used here is roughly similar to the Rayleigh criterion.

Use of the general equation in this manner enables one to predict the behavior of any particular tube design under different conditions of operation. Once the calculations are made, it is a very simple step to compute the advantages of designing a tube with additional stages.

To determine how closely the experimental behavior of the tube checks the calculated, the data shown in Figs. 5 and 6 were obtained using the mass 20 ion of neon.

The value for percent of maximum energy gain was

obtained experimentally by using the ratio of the stopping potential to that potential required to just stop all ions from reaching the collector. Figure 5 compares the calculated resolution change to that obtained experimentally as the stopping potential is increased. The experimental value was based on the Rayleigh criterion of resolving power; $M/\Delta M$, where ΔM is the peak width at the half-height. It was not possible to use this identical criterion for the calculated values. Therefore, one would expect only the shapes of the curves to be similar, which Fig. 5 clearly indicates.

Figure 6 compares the calculated sensitivity change as one increases the stopping potential to that obtained experimentally. Here again, absolute values could not be compared, only the shapes of the two curves. It is not surprising that the shape of experimental sensitivity curve does not agree more closely to the calculated one. It would be difficult for the calculations to take into account the unknown effects of ion generation, ion recombinations, and space charges.

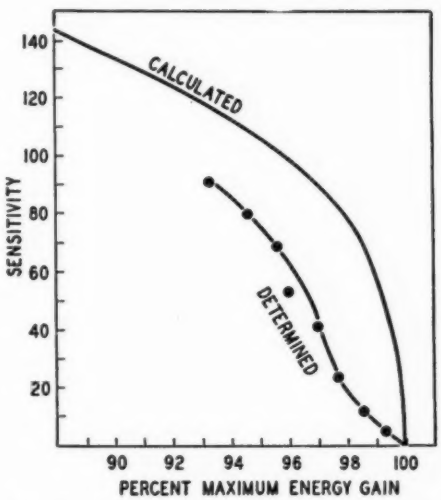


FIG. 6. Calculated *versus* determined sensitivity values at mass 20 ion.

III. PERFORMANCE

The mass spectra of hydrocarbons from methane through pentanes have been determined. Cracking patterns are quite similar to published patterns obtained on conventional magnetic deflection instruments. A linear mass scale which follows exactly the equation $M=0.192Va$ is obtained. Figure 7 shows the mass spectrum of *n*-butane, using an ionizing current of 200 microamperes and a pressure of 3×10^{-5} mm. The signal-to-noise ratio of the mass 43 ion in Fig. 7 is 2500/1.

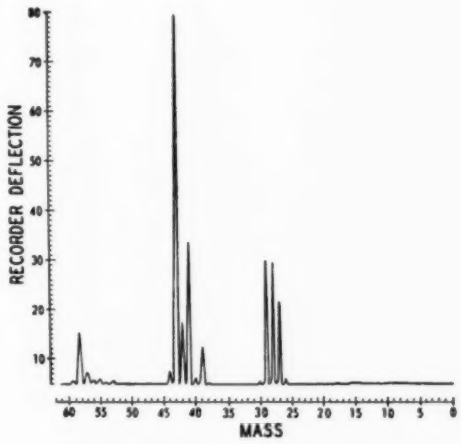


FIG. 7. Mass spectrum of *n*-butane.

Ion peak height stability of the instrument has been determined by cyclically scanning different mass peaks for 24 hour periods. A series of tests conducted over several months indicate an average peak height drift produced by instability in the tube and electronics combined to be around ±0.2 percent an hour.

The tube displays a marked indifference to ambient temperature changes. When the envelope temperature is varied from 40°C to 250°C, the change in the spectrum of *n*-butane may be attributed to change in temperature of ionization. Since no temperature control of the ionization is used, this is to be expected.

Numerous hydrocarbon mixtures with five and six components have been analyzed by conventional techniques. Most analytical results are within ±0.4 mole percent of the known values.

IV. CONCLUSIONS

While further work is required to determine conditions of optimum operation, the preliminary experimental data produced from a 5-stage nonmagnetic mass spectrometer tube are encouraging. They indicate the tube is capable of sufficient stability, resolving power, and sensitivity for use in the instrumentation of many analytical and research problems. The electronic circuits required to achieve a reasonable stability are relatively simple.

Silicon *n-p-n* Grown Junction Transistors

M. TANENBAUM, L. B. VALDES* E. BUEHLER, AND N. B. HANNAY,
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received January 29, 1955)

Silicon *n-p-n* junction transistors have been made from rate-grown single crystals. Using gallium and antimony as doping agents, single crystals of *n*-type silicon have been grown containing up to five *p*-regions from 0.0005 to 0.002 inches wide which are suitable for the production of transistors. The ohmic contact to the *p*-type base region was made by alloying an aluminum wire using techniques similar to those employed in the fabrication of aluminum-silicon diodes.

The electrical evaluation of *n-p-n* transistors produced from these single crystals is described. Alphas in excess of 0.9 at −1 ma emitter current have been obtained, and collector saturation currents of 10^{-6} ampere/cm² are common. With base-layer widths of about 0.0005 inch, the alpha cutoff occurs at approximately 5 megacycles. The units (approximately 0.04×0.04×0.5 inches in size) have been operated at power levels in excess of 1.5 watts in air with no special provision for heat dissipation.

INTRODUCTION

BECAUSE silicon has a larger forbidden energy gap than germanium (1.08 ev *vs* 0.72 ev), silicon devices inherently have at least two advantages over their germanium counterparts which are of great importance for many practical applications. First, silicon devices may be operated at much higher temperatures than germanium devices; temperatures up to 200°C are feasible. Secondly, the higher energy gap leads to reverse currents in silicon *p-n* junctions that are orders of magnitude lower than those in germanium junctions.

The silicon transistors which will be described here have been made from *n-p-n* material produced by the rate-growing technique. The growth of silicon single crystals is described in the first part of this paper. The techniques used for making the transistors and the effect of heat treating the silicon are discussed in the next section. Finally, the characteristics of these silicon *n-p-n* transistors are summarized.

1. GROWTH OF SINGLE CRYSTALS CONTAINING *n-p-n* JUNCTIONS

Thin layer *n-p-n* or *p-n-p* structures have been made in germanium by the double-doping technique[1] and by rate growing[2]; both of these techniques have been successfully applied to silicon. The rate-growing technique depends on the fact that in closely compensated melts, the conductivity type of the grown material is a sensitive function of growth rate. This fact makes it possible to produce thin layers of differing conductivity types in a growing crystal by inducing transients in the growth rate that are large compared to the small uncontrolled fluctuations which are always present in any practical crystal-growing machine.

The concentration distributions of antimony, boron, and gallium as a function of length in a silicon crystal grown at a constant rate are shown in Fig. 1. These curves are based on the latest available figures for the distribution coefficient, k, of these elements: that is $k_{Sb}=0.04$, $K_{Ga}=0.01$, and $k_B=0.9$. In order to produce rate-grown junctions, the initial average melt concentrations, C_0 were chosen so as to provide near-compensation for most of the length of the crystal. Although the exact dependence of k on the growth rate has not been determined in silicon, combinations of boron with antimony and of gallium with antimony have been employed successfully to produce *p*-layers in *n*-type crystals by introduction of a change in the growth rate. Experience with boron and antimony indicated that the concentration ratio would be useful over only a short distance. This is anticipated from Fig. 1 which indicates that the concentration ratio of the two elements in the melt will change radically as the crystal is pulled. Antimony and gallium were chosen for the doping impurities since their concentration-distribution curves are almost parallel and the concentration ratio in the melt should remain substantially constant.

The antimony was added after the silicon was

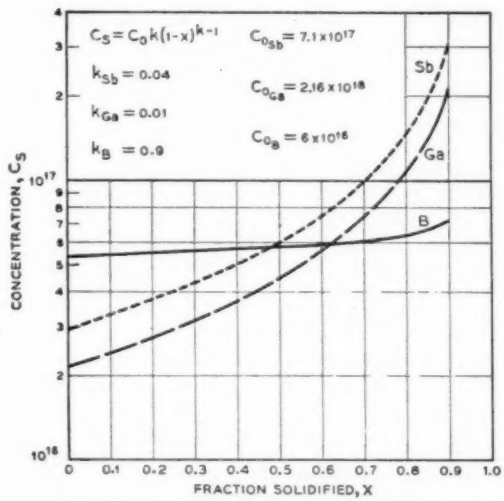


FIG. 1. Distribution of impurities in a pulled silicon crystal.

* Present address; Pacific Semiconductors, Inc., 10451 W. Jefferson Blvd., Culver City, California.

[1] Teal, Sparks, and Buehler, Phys. Rev. **81**, 637 (1951).

[2] R. N. Hall, Phys. Rev. **88**, 139 (1952); R. N. Hall, J. Phys. Chem. **57**, 836 (1953).

molten in order to avoid excess loss of antimony by vaproization. A single-crystal seed was then inserted into the melt and the crystal was grown by slowly withdrawing the seed from the melt. Junctions were grown into the crystal by causing the rate of crystallization to go from a positive value to a negative value (meltback) and then back to the original positive value. This was accomplished by changing both the temperature of the melt and the rate at which the seed was withdrawn. This cycle was repeated as many times as the volume of the melt would permit. An average crystal would contain five or six n-p-n structures.

The relative concentrations of Ga and Sb which one might expect to be produced in the crystal by such growth variations are shown in Fig. 2. The meltback should give an n-p step junction which is a good emitter. The second junction should be more graded and a good collector up to higher voltages. The electrical measurements to be reported in Sec. 3 substantiate this picture of the structure of the p-region.

Figure 3 is a photograph of an early crystal with five junctions. The p-type layers, which have been selectively darkened by etching, are thin in the center and become wider at the edges. Layers near the seed end are always thinner than those near the tip end since the concentration ratio of Ga/Sb in the melt does, in reality, increase as the crystal is withdrawn.

Because of certain heat-treating characteristics, which are discussed in more detail in Sec. 2, it was advantageous to grow n-p-n structures in crystals which were not rotated during growth.

In addition to these heat-treating characteristics, another difficulty is observed when rate-grown junctions are made in rotated crystals. Because of an uneven wetting of the quartz crucible by the molten silicon, the melt tends to "hang up" on certain sides of the crucible and thus presents an asymmetrical and nonplanar surface to the growing crystal. If the crystal is rotating, during the course of a single revolution a particular spot will at one moment see hot melt and at the next moment will see relatively cold gas ambient. Such variations produce rapid fluctuations in the growth rate, and in such a closely compensated crystal produce rate-grown p-layers where they are not wanted. Figure 4 is a photograph of a section of such a rotated crystal. An etch has been used which preferentially darkens the

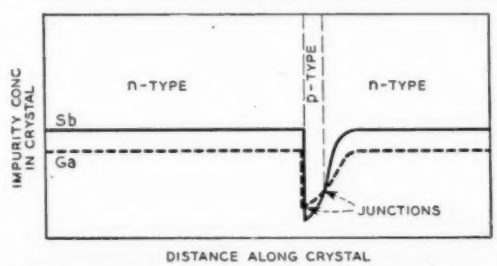


FIG. 2. Impurity concentrations in rate-grown silicon n-p-n junction transistors.

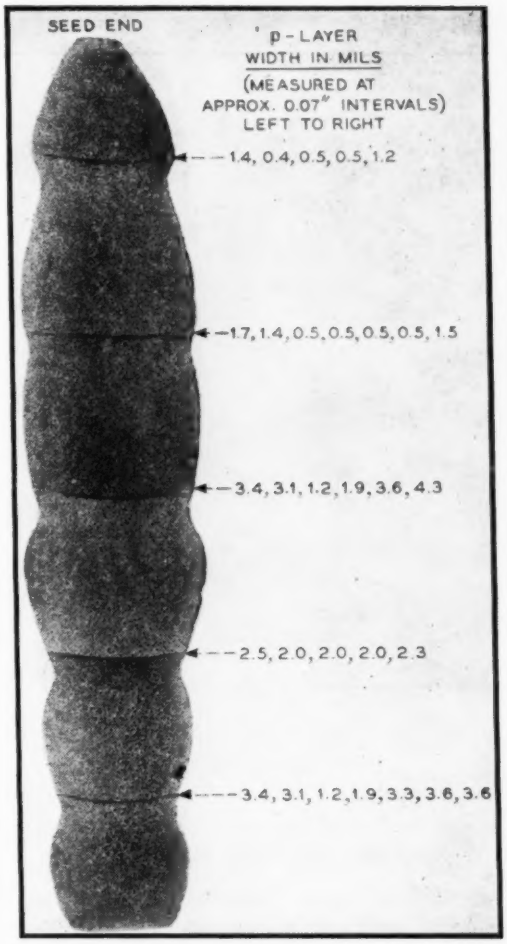


FIG. 3. Section of rate-grown silicon crystal.

p-type regions. The spacing between these p-regions corresponds exactly to the rate of rotation of the crystal during growth. These p-regions form a spiral around the crystal with pitch equal to the rate of rotation divided by the growth rate.

A rate-grown p-layer was produced in the section near the center of the photograph and the variations in growth rate are evidenced by the changes in spacing of the striations. The fact that this spacing is appreciably larger just below the p-layer than a short distance away on both sides of the layer shows that although the pull rate is returned to its original value after meltback, the actual growth rate is somewhat more rapid for a brief period and then decreases to its original value.

2. PREPARATION OF TRANSISTORS

Before describing the preparation of transistors from rate-grown n-p-n structures, it is necessary to review briefly the results of a detailed series of investigations by C. S. Fuller[3] concerning the effect of heating silicon single crystals at elevated temperatures.

[3] Fuller, Ditzenberger, Hannay, and Buehler, Phys. Rev. **96**, 833(A) (1954); Acta Metallurgica, **3**, 97 (1955).

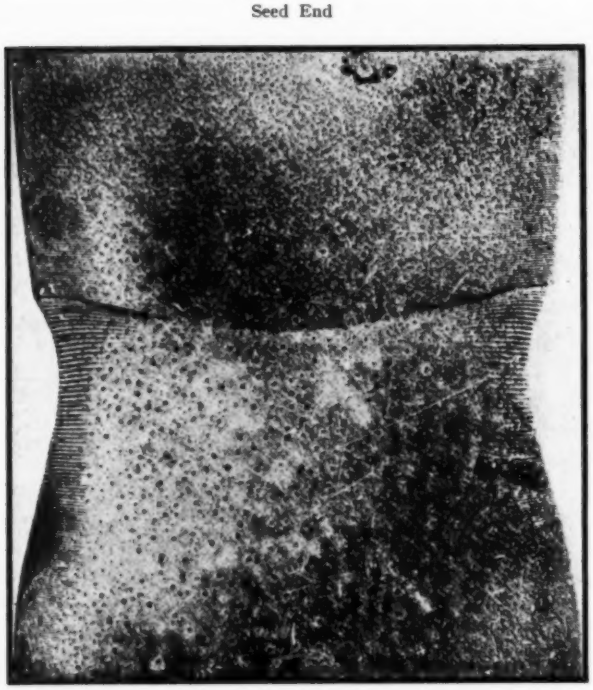


FIG. 4. Conductivity striations in a rotated silicon crystal.

Although considerable variability in the effect of heating different crystals was observed by Fuller, the following generalizations could be made:

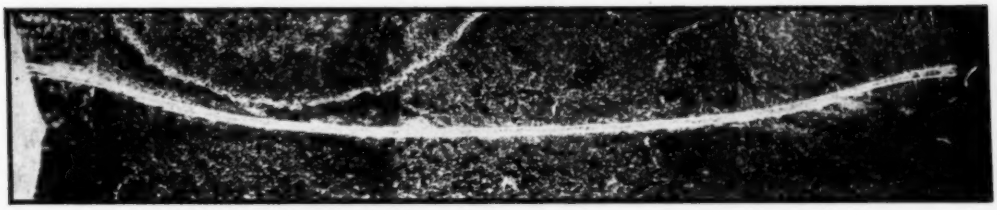
(1) Heating a crystal (grown with normal rotation) below 500°C results in an increase in N_D-N_A (conversion toward n-type), where N_D and N_A are the concentrations of donors and acceptors, respectively. This change may require from several minutes to several hours to reach a stationary state, but usually several hours are required.

(2) Heating above 500°C results in a decrease in N_D-N_A (conversion toward p-type). At 1000°C this change generally reaches a stationary state in a matter of seconds.

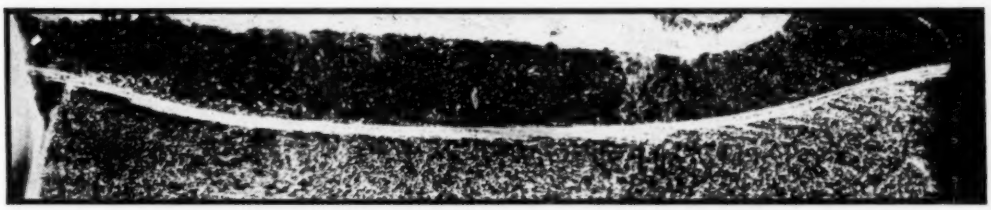
Fuller also noted that grown crystals of diameter less than one quarter of an inch did not show much change on heating. In the course of a study of crystal-growing variables by Buehler and Hannay, crystals were grown without rotation, and it was observed by Fuller that these also did not change in resistivity upon heating. The mechanisms of these changes in silicon produced by heating are not entirely understood at present, but with a knowledge of the experimental facts in mind it has been possible not only to circumvent the effect of heat treatment, but even to utilize the effect to produce structures which would be very difficult to prepare by conventional crystal-growing techniques.

The effect of heating is of importance in the present work because of the necessity of making contact to the very narrow base region in a grown-junction transistor structure. It is necessary to make a contact to this region that is mechanically strong and of low resistance. Since there is at present no mechanically feasible method of making a contact which meets these requirements to a base layer which is a thousandth of an inch or less in width without overlapping into the emitter or collector regions of the unit, the contact must also be rectifying to these end regions. It then becomes merely an extension of the base and little harm is done. The most convenient contact that is presently feasible and that meets the preceding specifications is a wire which is either alloyed or electrically bonded to the base layer. The wire must either be of a material which will dope silicon to the same conductivity type as the base layer or else consist of an inert material with a sufficient concentration of such a doping agent. Since aluminum is known to give low resistance contacts to p-type silicon and excellent rectifying contacts to n-type silicon, it was the obvious choice for the n-p-n structure. However, in the ideal case it is necessary to heat the silicon and the aluminum at least to the eutectic temperature (577°C) in order to make a satisfactory contact, and in actual practice higher temperatures are required. It might be expected that one would have to worry about heat-treatment effects, and indeed this was found to be the case.

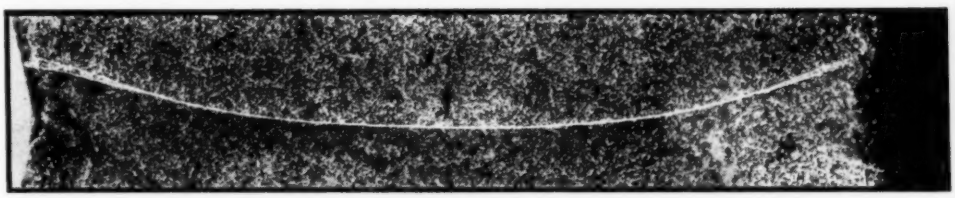
The effects of resistivity changes produced by heating the silicon during alloying are best shown by considering the p-layer of a typical crystal which was rotated during growth. Figure 5(a) shows the $BaTiO_3$ pattern of the p-layer obtained on a half of the crystal as received from the crystal puller. The base width is about 0.002 in. The crystal was then placed on a clean strip of tantalum metal which was heated to the alloying temperature by passing a large current through the strip in an atmosphere of tank helium. The crystal section was heated in this manner for several seconds until it reached peak temperature and was then allowed to cool rapidly in helium. This is the same technique that was generally used when alloying an aluminum contact to the base layer of a unit, the only differences being that no aluminum was present in this case, and a longer time was required to get the more massive crystal section up to the strip temperature. The $BaTiO_3$ pattern of the base layer after this heating is shown in Fig. 5(b). The layer has become somewhat wider and shows striations on the convex side of the layer. This is away from the seed and is thus the rate-grown junction. The remelt junction remains sharply defined. The widening is explainable if the actual structure of the junction is that which is postulated in Fig. 2, and if the heat treatment has served to decrease the relative value of N_D, i.e., decrease N_D-N_A. The striations may be explained if the increase in N_D near the rate-grown junction is not a smoothly increasing function of distance as shown in Fig. 2, but instead oscillates with small amplitude about this smooth curve. There is



(a) Base layer as received



(b) Base layer after alloying cycle



(c) Base layer after annealing at 475°C

FIG. 5. Heat-treatment effects in a rotated silicon crystal near a rate-grown p-layer.

ample evidence of such oscillations in more extreme cases to permit this assumption in this case.[4]

Thus, after alloying the base contact, the junction may be wider than in the as-grown condition and may also contain an undesired striated structure. ("Bonding" the aluminum wire, which consists of passing a large current through the aluminum wire and the cold specimen and thus produces the alloying temperature in a localized region, has been shown to produce similar alterations in the base region.) However, Fuller's results show that by annealing at 475°C, it is possible to return the silicon to a state of maximum N_D. Indeed, by controlling the time at 475°C, the base layer width may be reduced to almost any desired thickness. Figure 5(c) shows the $BaTiO_3$ pattern of the same crystal section after 16 hours at 475°C in air following the strip heating described above. End-to-end electrical measurement shows clearly that an n-p-n structure still exists, but both junctions cannot be resolved with $BaTiO_3$, i.e., the width of the p-region is less than 0.0005 in.

A bar cut adjacent to this half section of the crystal shown in Fig. 5 was given a similar annealing at 475°C after an aluminum wire was bonded to the p-layer. In this manner a silicon transistor with an alpha of 0.9 at -1 ma emitter current was produced.

Although the technique described above is a powerful method of producing base layer widths which would be difficult to produce by conventional techniques, there is the fundamental objection that a phenomenon is utilized which is not as yet fully understood, and therefore cannot be completely controlled. The experiments of Buehler, Fuller, and Hannay show, however, that if n-p-n structures are grown without rotating the crystal, there will be no resistivity changes during alloying. As reported in the foregoing, this has been done successfully and reproducibly and, as predicted, little change in the base layer structure was apparent either after alloying or after a succeeding heat treatment at 475°C.

The procedure that was finally adopted to prepare nonrotated rate-grown silicon transistors is outlined below. Bars 0.04×0.04 in. in cross section were cut from the crystal and chemically polished in an etching solution of hydrofluoric and nitric acids. A 60-cycle alternating voltage was applied to the ends of the specimen, and the p-layer was located by applying a dilute suspension of $BaTiO_3$ powder in carbon tetrachloride. An aluminum wire was then alloyed to the base layer by heating the specimen in an inert atmosphere. Contacts to the emitter and collector regions were made by sandblasting and rhodium plating the ends of the bars and then soldering directly to the rhodium plate. A final etch in the hydrofluoric-nitric acid mixture completed the preparation.

Although the etch after alloying and mounting was

[4] Burton, Kolb, Slichter, and Struthers, J. Chem. Phys. **21**, 1991–1996 (1953).

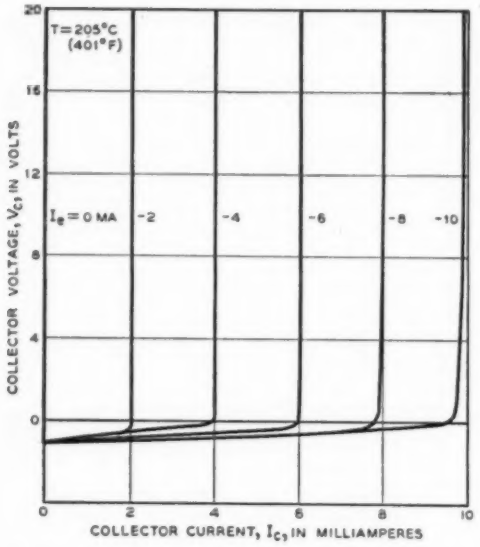


FIG. 6. Collector characteristics of a silicon *n-p-n* rate-grown transistor at elevated temperature. This figure is a tracing of an oscilloscope pattern.

necessary to remove the debris around the base layer formed during these operations, no further etching has been necessary. The general behavior of the units has not changed significantly after several months from the behavior immediately after etching although none of these transistors is hermetically sealed. In general, it is necessary to use *p*-layers less than 0.001 in. wide in order to obtain alphas greater than 0.9. This is probably because of the large surface recombination velocity on silicon which has not been reduced significantly. Since these thin layers result in small surface-to-volume ratios, it is not surprising that the alpha of these units is relatively stable over long periods of time.

3. ELECTRICAL CHARACTERISTICS OF SILICON TRANSISTORS

The temperature and power advantages that silicon devices would possess have been evident for some time. However, it has also been known that a critical evaluation of silicon transistors must wait until there are devices which obey the simple theory and can be treated analytically. The transistors which are described here, despite their relatively complex structure, seem to behave according to the analytical model for planar *n-p-n* transistors.

TABLE I. Comparison of silicon and germanium transistors.

	Silicon research model	A-1859 Development model
$h_{11}(-1, 4.5)$—ohms	154	200 max
$h_{12}(-1, 4.5)$	0.4×10^{-3}	10^{-3} max
$h_{21}(-1, 4.5)$	−0.919	−0.85 min
$h_{22}(-1, 4.5)$—micromhos	0.47	1.0 max
$I_c(0, 4.5)$—amperes	3.5×10^{-9}	10^{-5} max
$C_c(0, 4.5)$—$\mu\mu$f	30	25 max
f_α of α—megacycles/sec	5	0.5 min
Noise fig.—decibels	25	30 max
Power ratings		
Collector voltage	20 max	30 max
Collector current—ma	80 max	5 max
Power dissipation—mw	1500 max	50 max
Ambient temp—°C	200 max	60 max

The most important advantage of silicon transistors is their ability to operate at high temperatures. This opens a new field of applications for transistors, since germanium devices are restricted to a maximum ambient temperature around 70°C. The feasibility of high-temperature operation is best illustrated with the aid of Fig. 6, which shows the collector characteristics of a typical experimental silicon *n-p-n* transistor at 205°C. Even at this temperature the collector saturation current at 20 volts was only 56 microamperes. At room temperature the saturation current under the same bias conditions was 10^{-8} amperes or one microampere per square centimeter of junction area.

Many of the pertinent transistor characteristics[5] are outlined in Table I. A typical silicon transistor has been compared with the characteristics of an experimental, low-grade, grown-junction, germanium *n-p-n* type. This comparison can be justified on the basis that the silicon transistor is a research model and very little development has yet been done to improve its characteristics.

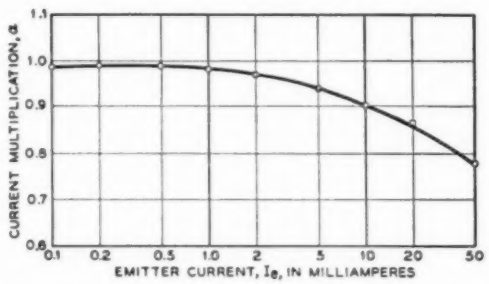


FIG. 7. Alpha as a function of emitter current in a silicon *n-p-n* rate-grown transistor.

The major component of the input impedance h_{11} is believed to be the ohmic resistance of the *n*-type material on the emitter side of the junction. The feedback[6] term h_{12} is primarily emitter voltage feedback μ_{ec} at small currents. When $I_e>1$ ma, the ohmic base resistance and the modulation of the base resistance produce feedback terms $r_b'h_{22}$ and μ_{22} which are comparable to μ_{ec}. The collector conductance h_{22} agrees with the values predicted by space-charge layer widening effects and increases linearly with emitter current.

The variation in α with emitter current is best

[5] The bias conditions, $I_e=-1$ ma and $V_c=4.5$ volts, are in parenthesis after the symbol for each parameter. Definition of the "*h*" parameter is found in E. A. Guillemin, *Communication Networks* (John Wiley and Sons, Inc., New York, 1931, 1935), Vol. II, Chap. IV.

[6] J. M. Early, Bell System Tech. J. 32, 1271–1312 (1954).

illustrated by Fig. 7, which shows data for another typical silicon transistor. Preliminary calculations show that the decrease in α at large currents[7] can be accounted for by a transverse drift field in the base region or a decrease in emitter efficiency caused by modulation of the conductivity of the base region.

The decrease in α at large currents is also visible in Fig. 8. These collector characteristics are intended to illustrate the possibility of high-power operation at room temperature. No special provisions have been taken to remove heat from the transistors. The bars, which are 0.04×0.04×0.5 in. are suspended by the three leads and sit in the center of an air-filled metal can (for mechanical protection).

The thin base region is desirable because it leads to a large alpha and high-frequency cutoff of alpha. However, it also increases the collector conductance h_{22}

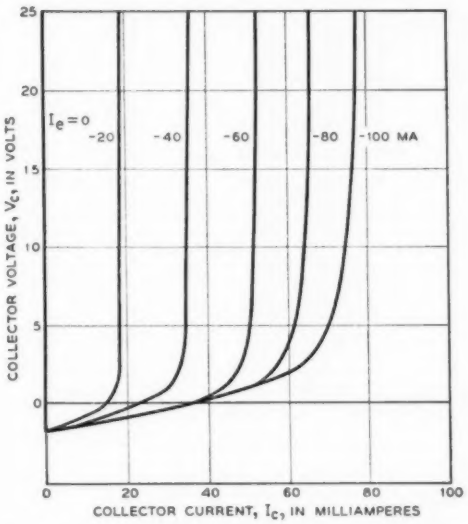


FIG. 8. High-power collector characteristics of a silicon n-p-n rate-grown transistor. This figure is a tracing of an oscilloscope pattern.

and the voltage feedback h_{12} at the emitter. The ohmic base resistance r_b' is also large because of the thin base region and the high resistivity caused by the structure of rate-grown material.

It has already been shown in Fig. 2 that the emitter junction is believed to be abrupt and the collector p-n junction to be graded. Further evidence to support this structure appears in a curve of the depletion layer capacitance of these transistors as a function of voltage. Figure 9 shows that the collector capacitance varies inversely as the cube root of the voltage. This is the behavior expected[8] from a p-n junction in which the impurity concentration increases linearly with distance on both sides. In the emitter the behavior is more complicated because there is a sharp rise in conductivity

[7] W. M. Webster, Proc. Inst. Radio Engrs. 42, 914–920 (1954).
[8] W. Shockley, Bell System Tech. J. 28, 435–489 (1949).

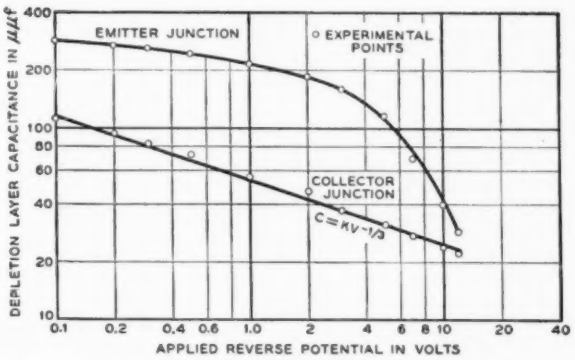


FIG. 9. Depletion layer capacitance of the emitter and collector junctions of a silicon n-p-n rate-grown transistor.

at the emitter step-junction which produces a small variation in capacitance with voltage. At larger voltages the space-charge region penetrates deeper into the base and into the region where there is a gradual decrease in conductivity on account of the graded collector junction. In this region the capacitance drops in magnitude very fast as the reverse emitter voltage is increased.

The noise figure of 25 db above thermal, which was indicated in Table I, was measured at 1000 cps with $I_e=-1$ ma and $V_c-4.5$ volts. The transistor was connected grounded-base with a generator resistance of 1000 ohms and a load resistance of 5000 ohms.

4. CONCLUSIONS

The transistors which are described here have been made from rate-grown n-p-n silicon crystals with no attempt to design for resistivities or to control the resistivity gradient in the thin base region. Essentially two basic technological improvements are reported here: (1) The ability to grow (nonrotated) silicon crystals with rate-grown n-p-n structures which do not show resistivity changes with heating; and (2) the possibility of utilizing to advantage in the fabrication of silicon transistors the heretofore undesirable effects of conductivity changes during heat treatment.

This paper shows that it is feasible to make silicon transistors from rate-grown structures. The total amount of minority carrier recombination at the surface has been reduced by making the base region very thin. This assures a fairly high alpha but also produces a large feedback voltage at the emitter. This large feedback, in itself, is the most undesirable feature of these transistors because the decrease in alpha with increase in emitter current can be tolerated in most cases except high-power applications. Increasing the base region width will reduce the feedback and the collector conductance. The ideal solution is to achieve complete control of the rate-growing process so that the structure and resistivity of the base region can be predetermined as easily as it is now possible to control

the base layer width. This will permit the design and preparation of transistors where the penetration of the collector space-charge region into the base region is minimized.

These transistors have the inherent advantages which come from utilizing silicon. In other words the saturation current of the collector is small and the transistors will operate at high temperatures. When either of these turns out to be a primary consideration, the transistors described here will perform a useful function.

ACKNOWLEDGMENTS

The authors are indebted to many people who have contributed to the development and characterization of these transistors. H. C. Montgomery made the noise-figure measurements. C. S. Fuller and J. A. Ditzenberger have helped by making their equipment and techniques available for heat treatment. J. D. Porcella aided in the construction of equipment and in the fabrication of the transistors. The encouragement and helpful suggestions given by M. Sparks have been invaluable.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 6 JUNE, 1955

Magnetic Materials for Digital Computer Components.* II. Magnetic Characteristics of Ultra-Thin Molybdenum-Permalloy Cores

NORMAN MENYUK
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts

(Received January 10, 1955)

Measurements of the flux-reversal time τ as a function of the applied magnetic field H have been made for ½-mil, ¼-mil, ½-mil, and 1-mil 4–79 molybdenum-Permalloy tape cores. There measurements have been taken at seven temperatures, ranging from −196°C to 270°C. From these measurements, the switching coefficient $S_w=(H-H_0)\tau$ is found, where H_0 is the threshold field value for irreversible domain-wall motion. Determination of the switching coefficient as the tape thickness is varied permits a separation of the spin-relaxation and eddy-current contributions to switching delay. These contributions are studied individually as functions of temperature. Upon increasing the temperature over the range considered, the eddy-current contribution and the threshold field value H_0 are approximately halved, and the spin-relaxation contribution is reduced by 20 percent. All these factors lead to a faster flux reversal at higher temperature for a given applied field. This behavior is in agreement with theoretical predictions.

I. INTRODUCTION

A RECENT study by members of this Laboratory has led to the conclusion that ultra-thin 4–79 molybdenum-Permalloy tape cores are superior to ferrites in switching circuits and stepping registers.[1] Among the factors favoring the metallic tapes are superior thermal properties, higher flux density, and faster flux-reversal characteristics.

In this paper the flux-reversal characteristics of ultra-thin 4–79 molybdenum-Permalloy tape cores are investigated as functions of tape thickness and temperature. A previous theoretical study relates these characteristics to eddy-current and spin-relaxation damping in polycrystalline ferromagnetic materials[2]; the study also predicts the changes to be expected when temperature and tape thickness are varied. These predictions are compared with the experimental results described below.

* The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

[1] Brown, Buck, and Menyuk, Trans. Am. Inst. Elec. Engrs. (Comm. and Elect.) **16**, 631 (1955).

[2] N. Menyuk and J. B. Goodenough, J. Appl. Phys. **26**, 8 (1955).

II. THEORETICAL CONSIDERATIONS

The switching time τ of ferromagnetic materials with square hysteresis loops is inversely proportional to the applied magnetic field. This can be expressed by the equation

$$(H-H_0)\tau=S_w, \tag{1}$$

where H is the applied magnetic field, H_0 is the threshold value for domain-wall motion, and S_w is a constant defined as the switching coefficient of the material.

The observed value of the switching coefficient is the sum of two independent effects. Thus

$$S_w=S_w{}^e+S_w{}^r, \tag{2}$$

where $S_w{}^e$ is the eddy-current contribution to the switching coefficient and $S_w{}^r$ is the spin-relaxation contribution. For ultra-thin metallic tapes, these contributions are related to the fundamental parameters of the material by the equations[2]

$$S_w{}^r=\frac{I_s\Lambda d}{(\gamma^2 I_s{}^2+\Lambda^2)\langle\cos\theta\rangle}\left(\frac{K}{A}\right)^{\frac{1}{2}}\approx\frac{\Lambda d}{\gamma^2 I_s\langle\cos\theta\rangle}\left(\frac{K}{A}\right)^{\frac{1}{2}} \tag{3a}$$

$$S_w{}^e=\frac{8\pi^2\alpha I_s r_m{}^2}{\rho c^2\langle\cos\theta\rangle^3}, \tag{3b}$$

FIG. 1. Input current pulse.

where I_s is the saturation magnetization, d is the maximum distance a domain wall moves during a flux reversal, Λ is the relaxation frequency, γ is the magneto-mechanical ratio, $\langle\cos\theta\rangle$ is the mean value of the cosine of the angle between the applied field and the direction of easy magnetization, K is the anisotropy constant, A is the exchange constant, r_m is one-half the tape thickness, ρ is the resistivity in statohm-cm, c is 3×10^{10} cm/sec, and α is a numerical factor $1<\alpha<2$ which is introduced to account for surface nucleation effects.[1]

As given in Eq. (3), $S_w{}^r$ is independent of tape thickness while $S_w{}^e$ is proportional to the square of the thickness. Therefore, any variation in S_w between metal-tape cores of different thickness is due only to the change in the eddy-current contribution. This fact affords an experimental means of separating $S_w{}^e$ and $S_w{}^r$ in order to study their properties independently.

III. EXPERIMENTAL RESULTS

This experiment was performed with 4–79 molybdenum-Permalloy tape cores with thicknesses of $\frac{1}{8}$-mil, $\frac{1}{4}$-mil, $\frac{1}{2}$-mil, and 1-mil, supplied by Magnetics Inc. In order that the cores might be as uniform as possible, all except the 1-mil tapes were prepared from the same melt; the 1-mil tapes are of different manufacture. Two cores of each thickness were used in this experiment; unless otherwise stated, all values given below represent the average value for both cores.

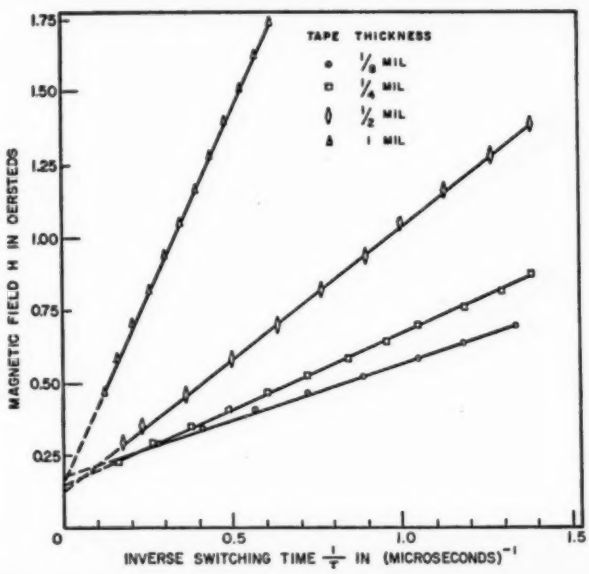


FIG. 2. Switching coefficient of tape wound 4–79 Mo-Permalloy cores at room temperature.

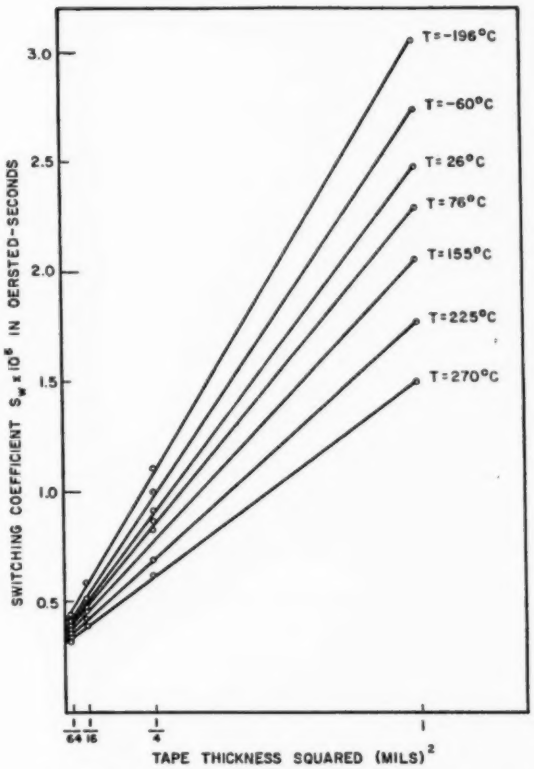


FIG. 3. Variation of switching coefficient with tape thickness as function of temperature.

A. Experimental Procedure

The input current used to reverse the magnetization of the cores is of the form shown in Fig. 1. It is effectively a constant-current signal with a rise time $t_r\approx0.2$ microsecond. The pulse length is long enough to permit a complete magnetization reversal of the core. The input-current amplitude was varied, and the switching time of each core was determined as a function of the applied magnetic field (based upon an average core diameter of 0.345 cm). This procedure was repeated for all cores at temperatures of −196°C, −60°C, 26°C, 76°C, 155°C, 225°C, and 270°C. A typical set of data obtained at room temperature (26°C) from a single core of each thickness is shown in Fig. 2.

The assumption of a constant applied magnetic field is not valid during the rise time of the pulse. This effect was minimized by limiting all measurements to field values for which the switching time is greater than, or equal to, 0.7 microsecond. The applied field is then constant during most of the switching cycle.

According to Eq. (1),

$$H=S_w/\tau+H_0. \tag{4}$$

This is the equation of a straight line of slope S_w and

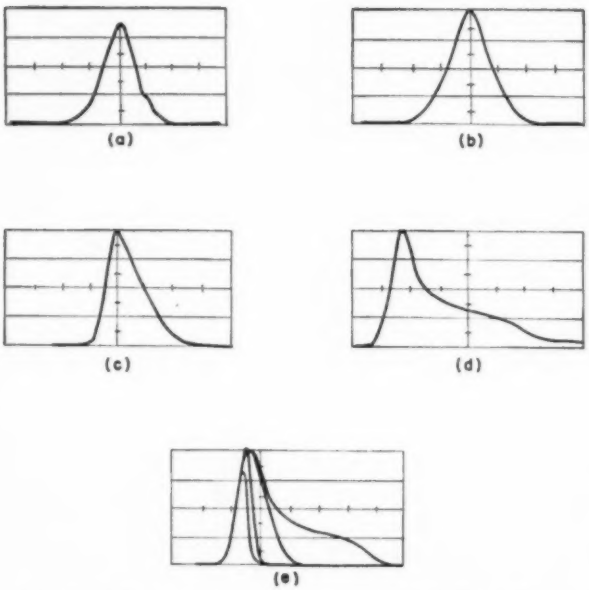


FIG. 4. Voltage output curves of 4–79 Mo-Permalloy tape cores. (a) $\frac{1}{8}$-mil core output; Time scale: 0.1 μsec/Div; Voltage scale: 135 mv/Div. (b) $\frac{1}{4}$-mil core output; Time scale: 0.1 μsec/Div.; Voltage scale: 155 mv/div. (c) $\frac{1}{2}$-mil core output; Time scale: 0.2 μsec/Div.; Voltage scale: 150 mv/Div. (d) 1-mil core output; Time scale: 0.3 μsec/Div.; Voltage scale: 150 mv/Div. (e) Superimposed output of $\frac{1}{8}$-mil, $\frac{1}{4}$-mil, $\frac{1}{2}$-mil, and 1-mil cores; Time scale: 0.3 μsec/Div.; Voltage scale: 150 mv/Div.

intercept H_0. The values of S_w and H_0 for each core are therefore directly obtainable from curves similar to those shown in Fig. 2.

B. Correlation of Theory and Experiment

Equation (2) and Eq. (3) lead to the relationship

$$S_w = S_w{}^r + D(2r_m)^2, \tag{5}$$

where

$$D = \frac{2\pi^2 I_s \alpha}{\rho c^2 \langle \cos\theta \rangle^3} \tag{5a}$$

is the eddy-current contribution to the switching coefficient per unit thickness, in units of oersted-seconds per centimeter squared. Since D is a constant at constant temperature, a plot of the switching coefficient *versus* tape thickness squared should yield a straight line of slope D and intercept $S_w{}^r$. A series of such curves, taken at different temperatures, is shown in Fig. 3. Since the curves are in units of oersted-seconds per mil squared, a conversion factor of 6.45 $\times 10^{-6}$ must be introduced. In all cases, the linear relationship holds well within the 10 percent thickness tolerance set by the manufacturer of the cores.

Since the intercept of the curves on the S_w axis in Fig. 3 represents the spin-relaxation contribution to the switching coefficient, it is apparent that the eddy-current contribution has been reduced to almost negligible proportions in the $\frac{1}{8}$-mil tape and is still small in the $\frac{1}{4}$-mil tape. In the $\frac{1}{2}$-mil tape the eddy-current contribution is comparable to the spin-relaxation contribution; in the 1-mil core the switching delay is predominantly an eddy-current effect.

Corroboration of these conclusions can be obtained from the shape of the output-voltage of the cores, as seen on an oscilloscope. Figure 4 is an oscilloscope tracing of the output-voltage signals obtained at room temperature upon reversing the magnetization of $\frac{1}{8}$-mil, $\frac{1}{4}$-mil, $\frac{1}{2}$-mil, and 1-mil 4–79 molybdenum-Permalloy cores with a magnetic field of 1.75 oersteds. Figure 5(a) shows a typical output-voltage signal obtained upon reversing the magnetization of a high-resistivity ferrite core in which eddy-current effects are negligible. The effect of eddy currents upon the shape of the output-voltage curve of a ferromagnetic tape has been theoretically determined by A. Papoulis[3] [Fig. 5(b)].

TABLE I. Variation of $S_w{}^e$, electrical resistivity and flux density with temperature.

Temperature (°C)	D (oe-sec/mil²)	ρ (statohm-cm)	$\alpha I_s/\langle\cos\theta\rangle^3$ (emu)
−196	2.68×10^{-6}		
−60	2.38×10^{-6}		
26	2.12×10^{-6}	6.11×10^{-17}	915
76	1.94×10^{-6}	6.45×10^{-17}	880
155	1.72×10^{-6}	7.00×10^{-17}	850
225	1.48×10^{-6}	7.45×10^{-17}	780
270	1.17×10^{-6}	7.75×10^{-17}	645

Comparison of Fig. 4 and Fig. 5 shows that the voltage-output curves of the $\frac{1}{8}$-mil and $\frac{1}{4}$-mil cores closely approximate that of the ferrite, where damping is almost entirely due to spin-relaxation effects. On the other hand, the 1-mil-core output approaches the shape of Fig. 5(b), indicating that eddy-current damping predominates. The $\frac{1}{2}$-mil-core output is somewhere between these extremes. Thus Fig. 4 and Fig. 5 qualitatively corroborate the quantitative results shown in Fig. 3.

The saturation magnetization and electrical resistivity are the only temperature-dependent variables on the right-hand side of Eq. (3b). Therefore, $S_w{}^e$ varies with temperature as I_s/ρ. The electrical resistivity of 4–79 molybdenum-Permalloy is given by Littman[4] as 55 microhm-cm at room temperature (6.1×10^{-17}

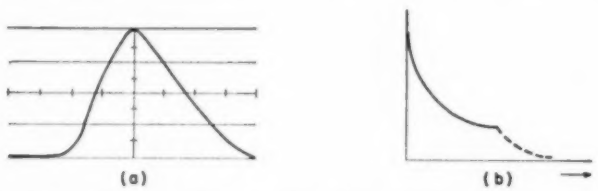


FIG. 5. Shape of voltage outputs due to relaxation and eddy-current effects. (a) Voltage output of a high resistance ferrite core in which eddy-current effects are negligible. (b) Predicted voltage output of a core driven by a constant current generator, considering eddy-current effects.

[3] A. Papoulis, J. Appl. Phys. **25**, 169 (1954).

[4] M. F. Littman, Trans. Am. Inst. Elec. Engrs. **71**, Part I, 220 (1952).

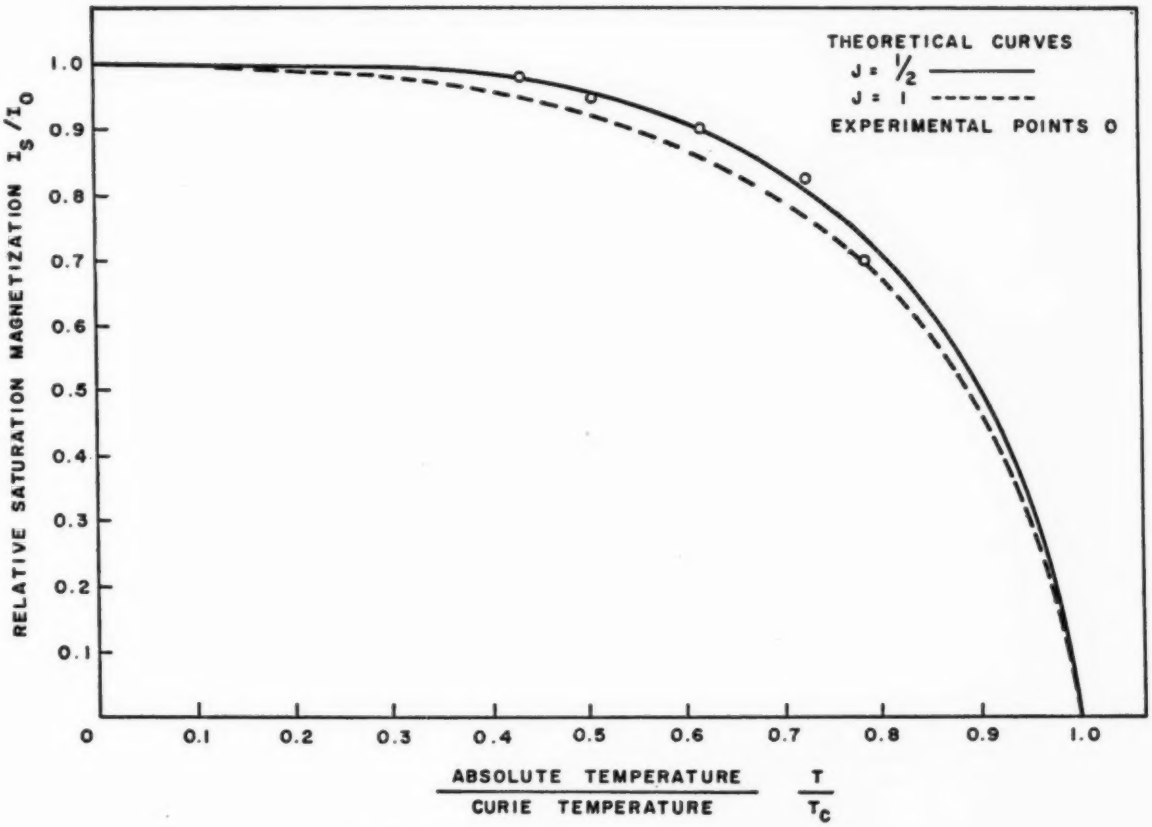


FIG. 6. Variation of magnetization with temperature.

statohm-cm), and the temperature coefficient of electrical resistivity is given by the Magnetic Metals Company as 0.0011 per degree centigrade in the range of 20°C to 500°C.[5] Within this temperature range, all the terms on the right-hand side of Eq. (5a) are known except I_s, α, and $\langle\cos\theta\rangle$. The experimentally determined value of D can therefore be used to derive $\alpha I_s/\langle\cos\theta\rangle^3$ as a function of temperature. The results are given in Table I.

There are no directly determined experimental values of I_s *versus* temperature available for comparison with these figures. However, since α and the angle θ are relatively temperature independent within the range considered, the experimental variation of I_s with temperature can be compared with the temperature variation predicted by the Brillouin function

$$\frac{I_s}{I_0}=\frac{2J+1}{2J}\coth\frac{(2J+1)a}{2J}-\frac{1}{2J}\coth\frac{a}{2J}$$

where

$$a=\frac{Jg\beta(H+NI_s)}{kT},$$

J is the angular-momentum quantum number, NI_s is the molecular field, β is a Bohr magneton, g is the Landé factor, k is the Boltzmann constant, I_s is the temperature-dependent saturation magnetization, and I_0 is the saturation magnetization at 0°K.

The normalized magnetization curve is shown in Fig. 6 for $J=\frac{1}{2}$ and $J=1$. The normalized experimental points are also shown for comparison. These points are based on the assumed value $\alpha I_0/\langle\cos\theta\rangle^3=940$ emu and on the Curie temperature $T_c=420$°C.[4] The close agreement of the theoretical and calculated experimental values strongly corroborates the prediction that $S_w{}^e$ varies as I_s/ρ.

The value of flux density B_s at room temperature is given as 8700 gauss. Assuming $B_s\approx 4\pi I_s$, this represents a saturation magnetization of 690 emu. The value given in Table I for $\alpha I_s/\langle\cos\theta\rangle^3$ is 915 emu. Thus $\alpha/\langle\cos\theta\rangle^3=1.3$. Since the molybdenum-Permalloy tapes are grain oriented, $\langle\cos\theta\rangle\approx 1$ and $\alpha\approx 1.3$.

The spin-relaxation contributions to the switching coefficient $S_w{}^r$ are obtained from the intercept of the curves in Fig. 3. The results, shown in Fig. 7, indicate a slow decrease of $S_w{}^r$ with increasing temperature over the range considered, changing from 0.4×10^{-6} oe-sec at −196°C to 0.32×10^{-6} oe-sec at 270°C. No anomalous effects occur within this temperature range.

[5] This figure is given for Hymu "80," the trade name of the Carpenter Steel Company for a material with the same composition as 4–79 molybdenum-Permalloy.

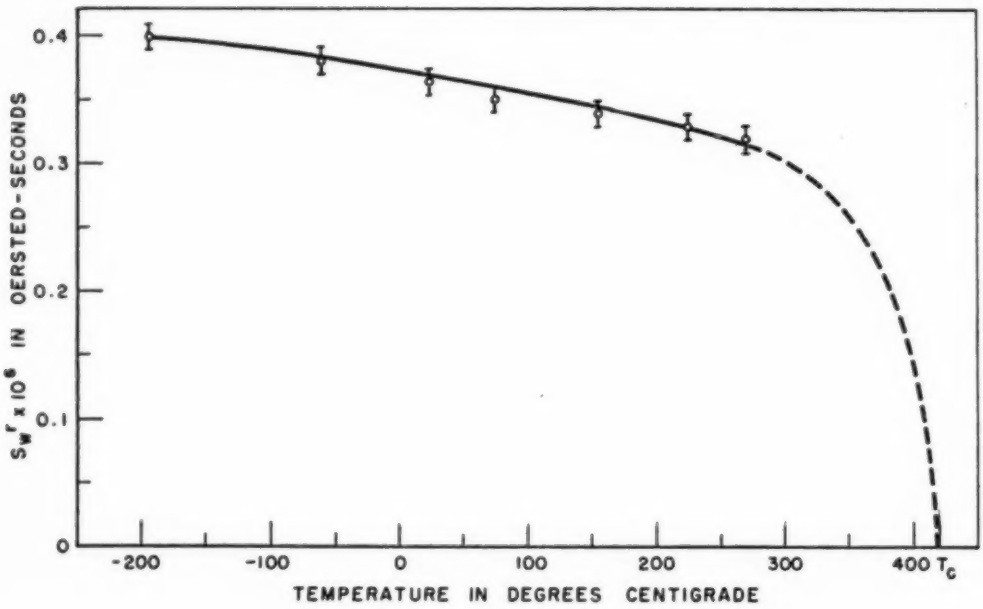


FIG. 7. Variation of S_w^r with temperature.

According to Eq. (3a)

$$S_w^r \propto \frac{\Lambda(K)^{\frac{1}{2}}}{I_s},$$

where all the other terms on the right hand are temperature independent. Actually, there should be a variation in d, but this variation is assumed to be small. The relative change of I_s with temperature is obtainable from Fig. 6. The temperature dependence of the relaxation frequency is not known for 4–79 molybdenum-Permalloy. However, Bloembergen[6] has investigated the relaxation frequency of Supermalloy, a material of similar composition, as a function of temperature. Bloembergen's results indicate that Λ increases by approximately 40 percent from room temperature to 270°C. Thus the variations of Λ and I_s are in the direction tending to increase S_w^r with temperature.

In order to overcome this tendency and to yield the decrease in S_w^r actually observed, there must be a fivefold decrease in the anisotropy constant K over this temperature range. Data given by McKeehan[7] shows that the anisotropy constant of an iron-nickel alloy containing 30 percent iron and 70 percent nickel decreases by a factor of 3.5 on going from 20°C to 200°C. Thus the figures obtained appear reasonable. Furthermore, the sharp decrease that must occur in S_w^r above 270°C to obtain $S_w^r=0$ at $T=T_c$ is also reasonable since (according to Van Vleck[8]) the anisotropy constant may be expected to vary rapidly in the vicinity of the Curie temperature.

The evidence cited above indicates experimental agreement with the predicted behavior. However, conclusive proof must await independent measurements of the values of Λ and K for 4–79 molybdenum-Permalloy as functions of temperature.

The threshold field H_0 represents the effective zero level which must be overcome for irreversible domain-wall motion. It is closely related to, but not identical with, the coercive force H_c. The coercive force of ultra-thin metal tapes decreases with increasing thickness.[4] This is also true of the threshold field values of the $\frac{1}{8}$-mil, $\frac{1}{4}$-mil, and $\frac{1}{2}$-mil tapes investigated in this experiment (which were all obtained from the same melt). The 1-mil tape has a threshold field value approximately equal to that of the $\frac{1}{4}$-mil tape. However, since the 1-mil tape was manufactured by a different company, variations in processing techniques may explain this apparent discrepancy. In addition, the threshold field decreases with increasing temperature which is in accord with the general behavior of the coercive force. The variation of the threshold field with temperature is shown in Fig. 8.

IV. CONCLUSIONS

The use of $\frac{1}{8}$-mil 4–79 molybdenum-Permalloy tape cores successfully reduces the eddy-current damping to less than 10 percent of the spin-relaxation damping at room temperature. Further reduction of tape thickness will therefore have a negligible effect upon the switching coefficient S_w. However, in tapes of $\frac{1}{2}$-mil thickness or greater, the eddy-current effect is predominant, and S_w increases rapidly with increasing thickness. This limits the use of 4–79 molybdenum-Permalloy to tapes of $\frac{1}{4}$-mil and $\frac{1}{8}$-mil for high-speed applications.

Upon increasing the temperature from −196°C to 270°C, the relaxation damping is decreased by 20 percent while eddy-current damping and the threshold

[6] W. Bloembergen, Phys. Rev. **78**, 572 (1950).
[7] L. W. McKeehan, Phys. Rev. **51**, 136 (1937).
[8] J. H. Van Vleck, Phys. Rev. **52**, 1178 (1937).

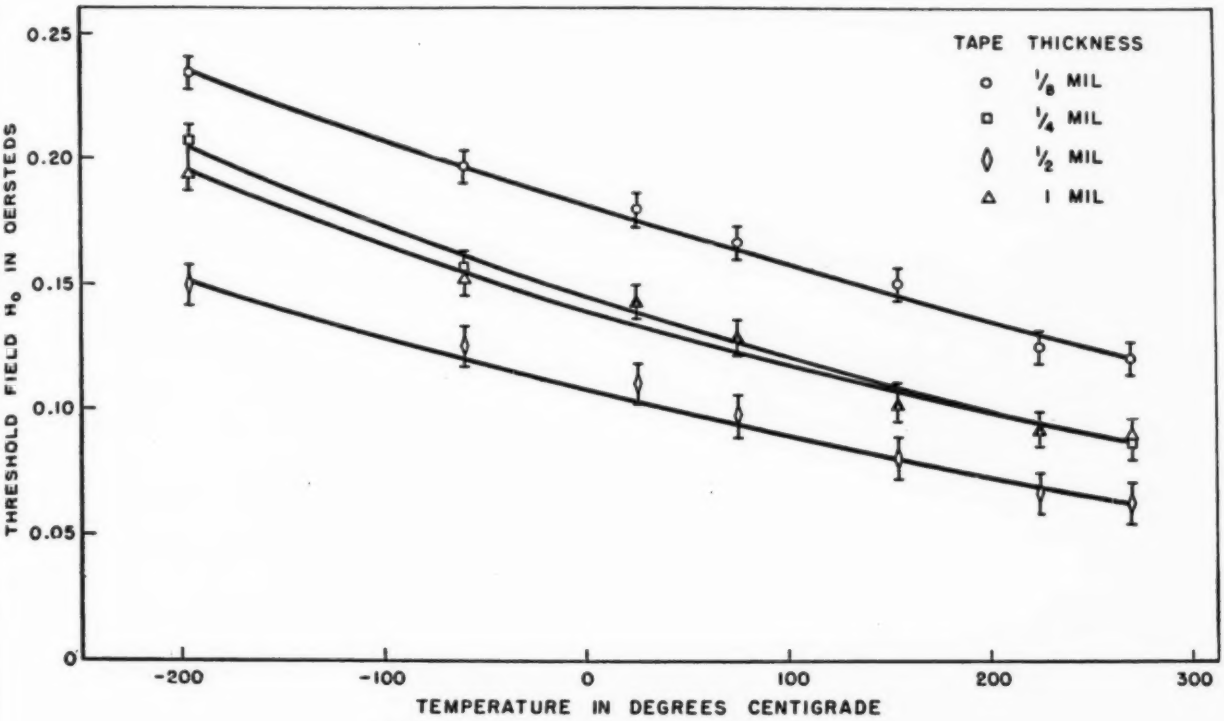


FIG. 8. Variation of threshold field with temperature.

field are approximately halved. Thus, for a given applied magnetic field, the switching time varies by a factor of approximately two over the temperature range. Although this variation of switching time with temperature is significant, it is too small to warrant the use of higher temperatures to attain faster core switching. On the other hand, the results indicate that the switching time is relatively insensitive to small temperature variations. For a $\frac{1}{8}$-mil tape core, a change of 20 degrees centigrade in the vicinity of room temperature should produce a switching-time variation of about 3 percent. For thicker tapes, temperature sensitivity is slightly higher.

The effects of tape thickness and temperature upon the eddy-current contribution to the switching coefficient $S_w{}^e$ strongly corroborate the theory discussed in Sec. II. The variation of $S_w{}^r$ with temperature also agrees with the theory within the limits of independent data presently available.

ACKNOWLEDGMENTS

The writer wishes to thank P. Fergus and H. Mogensen for their help in obtaining the experimental data and wishes also to thank J. Childress for his assistance in setting up the experimental equipment.

Transition Through a Contact Region

J. G. HALL
Institute of Aerophysics, University of Toronto, Toronto, Ontario
(Received April 27, 1954; revised manuscript received January 17, 1955)

The one-dimensional contact region common to shock-tube flows is idealized by the simple model of initial step change in temperature and density at constant pressure. For weak transitions, neglect of pressure effects, and linearization of the conservation equations gives spatially monotonic temperature and density profiles with the contact region thickening as the square root of time. Experimental results are noted from shock-tube studies of two air/air contact regions. Insofar as comparison of theory and experiment is possible, the theory for weak transitions appears reasonable.

INTRODUCTION

THE inviscid one-dimensional shock-tube theory[1,2] represents the contact surface (region) as a plane discontinuity in density and temperature at constant pressure and flow velocity. In reality such a discontinuity cannot exist. Since the contact region moves with the fluid and always consists of essentially the same fluid particles, heat transfer must destroy any steep temperature gradient initially present. The present note gives a simplified analysis of the physical transition. Experimental shock-tube results are also described and compared with theory where possible.[3]

THEORY

A homogeneous or single gas throughout is considered. The phenomenon is idealized by an initial step change in density ρ and absolute temperature θ at constant pressure p located at the origin of a reference frame moving with the gas at constant flow speed. As illustrated in Fig. 1, this gives the following one-dimensional initial value problem for $T(x,t)=\theta(x,t)/\theta_0$.

$$T(x,0)=\theta_1/\theta_0=T_1 \quad \text{for} \quad 0<x\leq\infty$$
$$T(x,0)=\theta_2/\theta_0=T_2 \quad \text{for} \quad -\infty\leq x<0$$
$$p_1=p_2$$
$$\theta_0=\tfrac{1}{2}(\theta_1+\theta_2).$$

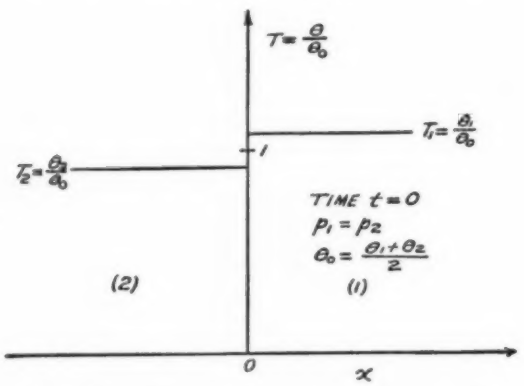


FIG. 1. Assumed initial conditions for contact region.

For consistency with this simple model and from pressure measurements[4,5] showing slight if any ordered change in static pressure through the transition region, the pressure p is taken as constant for $t>0$. Assuming an ideal gas, the state equation may then be written $T\Gamma=1$, where $\Gamma=\rho/\rho_0$. It might be noted that while the transition is essentially heat conduction at constant pressure, the existence of a density time derivative implies a mass velocity $u(x,t)$ in order that mass be conserved.

A solution may be sought from the continuum equations expressing conservation of mass, momentum, and energy. Neglecting body forces and bulk or second viscosity, the previous assumptions reduce the equations as follows:

$$\text{mass} \quad \frac{\partial\rho}{\partial t}+\frac{\partial}{\partial x}(\rho u)=0 \tag{1}$$

$$\text{momentum} \quad \rho\left(\frac{\partial u}{\partial t}+\frac{u\partial u}{\partial x}\right)=4/3\frac{\partial}{\partial x}\left(\mu\frac{\partial u}{\partial x}\right) \tag{2}$$

$$\text{energy} \quad \rho c_v\left(\frac{\partial\theta}{\partial t}+\frac{u\partial\theta}{\partial x}\right)$$
$$=\frac{\partial}{\partial x}\left(k\frac{\partial\theta}{\partial x}\right)-p\frac{\partial u}{\partial x}+\frac{4}{3}\mu\left(\frac{\partial u}{\partial x}\right)^2 \tag{3}$$

[1] Glass, Martin, and Patterson, "A theoretical and experimental study of the shock tube," UTIA Report No. 2, 1953. Institute of Aerophysics, University of Toronto.

[2] J. Lukasiewicz, "Shock tube theory and applications," N.A.E. Report 15, 1952. National Aeronautical Establishment, Ottawa.

[3] Further details of the work than given here may be found in "The transition through a contact region," J. G. Hall, UTIA Report No. 26, 1954. Institute of Aerophysics, University of Toronto.

[4] Huber, Fitton, and Delpino, "Investigation of moving pressure disturbance and correlation with one-dimensional flow theory," NACA Technical Note No. 1903, 1949.

[5] E. K. Parks made piezo-crystal pressure gauge and schlieren studies of the diffraction of a plane shock wave about a sharp 30° convex corner as well as of similar diffraction for a plane shock moving along a constant area duct into a sharp-cornered two-dimensional Busemann nozzle. His experimental results showed that a contact region or entropy transition was formed immediately following the diffracted incident wave. See "Supersonic flow in a shock tube of divergent cross-section," E. K. Parks, UTIA Report No. 18, 1952. Institute of Aerophysics, University of Toronto.

where μ = shear viscosity, k = heat conduction coefficient, and c_v = specific heat at constant volume. The neglect of pressure change leaves only three dependent variables whereas there are four equations to be satisfied. However, approximate solutions may be obtained which satisfy state and the linearized forms of (1), (2), and (3) for a Prandtl number $\sigma=\mu c_p/k$ of 3/4.

For a weak initial temperature step, i.e. $T_1-T_2\ll 1$. Eqs. (1) to (3) may be linearized on the assumption that mass velocity u and all derivatives are small. Assuming a viscosity law $\mu=\mu_0 T$ and constant Prandtl number, one obtains the following linear approximations to (1), (2), and (3) by neglecting quadratic and higher order terms.

$$\frac{\partial T}{\partial t}=\frac{\partial u}{\partial x} \tag{4}$$

$$\frac{\partial u}{\partial t}=\frac{4\mu_0}{3\rho_0}\frac{\partial^2 u}{\partial x^2} \tag{5}$$

$$\frac{\partial T}{\partial t}=\frac{\mu_0}{\sigma\rho_0}\frac{\partial^2 T}{\partial x^2}. \tag{6}$$

The linearization will not be valid for $t\ll l_0/c_0$ where l_0 = mean free path and c_0 = average molecular speed.

The solution for the energy Eq. (6) which satisfies the initial conditions is

$$T=1+\tfrac{1}{2}(T_1-T_2)\,\mathrm{erf}\left(\frac{x}{2(At)^{\frac{1}{2}}}\right) \tag{7}$$

where

$$A=\frac{\mu_0}{\sigma\rho_0} \quad \text{and} \quad \mathrm{erf}z=\frac{2}{\pi^{\frac{1}{2}}}\int_0^z \exp(-s^2)ds.$$

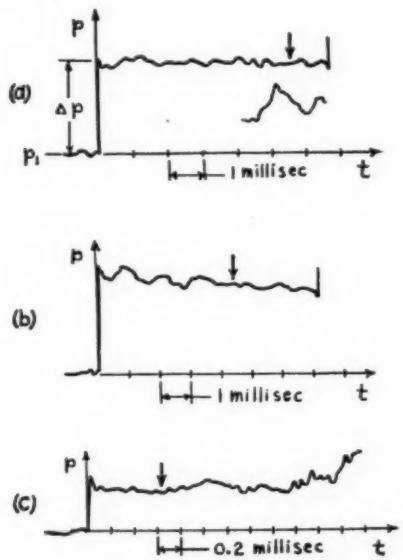


FIG. 2. Piezo-crystal pressure recordings through type A contact region. (a) Diaphragm pressure ratio $P=4$; $\Delta p=174$ mm Hg; $p_1=185.7$ mm Hg. (b) $P=8$; $\Delta p=148$ mm Hg; $p_1=93.2$ mm Hg. (c) $P=1024$; $\Delta p=31.0$ mm Hg; $p_1=3.0$ mm Hg. Arrow denotes beginning of contact region.

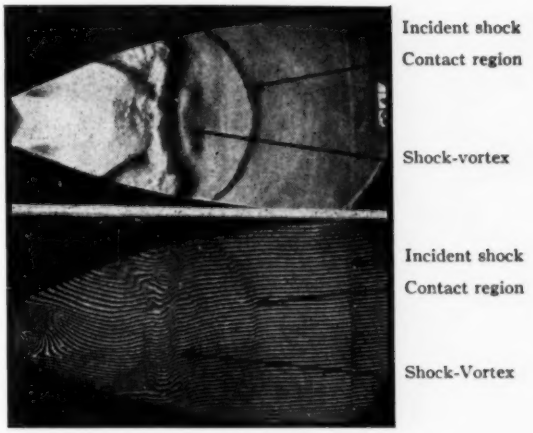


FIG. 3. Schlieren photo and interferogram showing development of type B contact region. Initial shock wave Mach number 2.20.

The mass velocity u is calculated from (4) and (7) as

$$u=\tfrac{1}{2}(T_1-T_2)\left(\frac{A}{\pi t}\right)^{\frac{1}{2}}\exp\left(-\frac{x^2}{4At}\right)+f(t) \tag{8}$$

which is seen to satisfy (5) exactly for $f(t)=0$ and $\sigma=\frac{3}{4}$. The linearized solution for density is obtained from state and (7) as

$$\Gamma=1+\tfrac{1}{2}(\Gamma_1-\Gamma_2)\,\mathrm{erf}\left(\frac{x}{2(At)^{\frac{1}{2}}}\right). \tag{9}$$

The linearization gives T and Γ distributions antisymmetric about the mean initial values, while the mass velocity u is symmetric about $x=0$. The assumptions for linearization, namely that u and all derivatives are small, are found *a posteriori* to be fulfilled in a time of order l_0/c_0.

The thickness $s(t)$ of the density transition may be defined as

$$s(t)=\frac{\Gamma_1-\Gamma_2}{h_0(t)}$$

where $h_0(t)=[\partial\Gamma/\partial x]_{x=0}$. Calculation from (9) then gives

$$s(t)=2(\pi At)^{\frac{1}{2}}. \tag{10}$$

The transition region thus thickens as the square root of time at a rate which increases with increase in kinematic viscosity μ_0/ρ_0.

An approximate solution which exhibits nonlinear effects might also be noted. The von Mises transformation

$$\frac{\partial\psi}{\partial t}=-\frac{\rho}{\rho_0}u$$

$$\frac{\partial\psi}{\partial x}=\frac{\rho}{\rho_0}$$

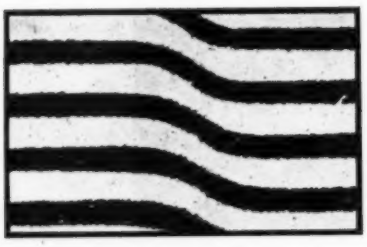


FIG. 4. Enlargement of fringe shift through central portion of type B contact region ten inches downstream from nozzle throat.

and the previous linear viscosity-temperature relation are used to change to independent variables $\psi(x,t)$, t in Eqs. (1) to (3). Linearization in the transformed plane leads to the following solutions[3] for $\sigma=\frac{3}{4}$.

$$T=1+\tfrac{1}{2}(T_1-T_2)\,\mathrm{erf}\left(\frac{\psi}{2(At)^{\frac{1}{3}}}\right) \tag{11}$$

$$u=\tfrac{1}{2}(T_1-T_2)\left(\frac{A}{\pi t}\right)^{\frac{1}{3}}\exp\left(-\frac{\psi^2}{4At}\right). \tag{12}$$

The transformation $x=x(\psi,t)$ is obtained as

$$x=\psi+(T_1-T_2)(At)^{\frac{1}{3}}\left\{\frac{1}{\pi^{\frac{1}{3}}}[\exp(-\eta^2)-1]+\eta\,\mathrm{erf}\eta\right\} \tag{13}$$

where $\eta=\psi/2(At)^{\frac{1}{3}}$. Although T and u possess symmetry when plotted against ψ, this symmetry is not retained on transforming back to the physical plane via (13).

EXPERIMENT

A 2 inch × 7 inch shock tube having interferometric, schlieren, and piezo-crystal pressure gauge instrumentation was used to study the following air/air contact regions:

A—That generated at the diaphragm station on initiation of shock-tube flow.

B—That generated by diffraction of a plane shock wave which propagated along a constant area duct into a sharp-corner-entry two-dimensional Busemann nozzle.[5]

Interferometric and schlieren results for the type A contact region showed a rough turbulent mixing zone unsuited for density transition or growth measurements—the mixing zone extended over several tube diameters and the density increased irregularly throughout. Because of diaphragm effects and boundary layer the contact front was strongly convex towards the downstream side.

Pressure recordings for diaphragm pressure ratios of 4 to 3000 showed the mean static pressure to be sensibly constant through the transition region. Figure 2 shows typical oscilloscope traces of the piezo-gauge response for diaphragm pressure ratios P of 4, 8, and 1024 at distances 9, 6, and 6 feet, respectively, from the diaphragm station. The arrow indicates the contact front as determined from schlieren photographs, and Δp is the pressure rise across the shock wave.

Type B contact region was limited to very small over-all density change with maximum interference fringe shifts of about one fringe spacing. The schlieren photo and interferogram of Fig. 3 show the development of the contact region behind the diffracted shock wave. Figure 4 is an enlargement through the central portion which was reasonably plane. Fringe shift profiles through the central portions were obtained from intensified plate enlargements with the aid of an optical comparator. Corrections for boundary layer and refraction were not considered.

Unfortunately, because of the spread of turbulence, the type B transition was too short-lived to measure the rate of thickening. Comparison of a particular density profile with theory is difficult as theory requires a time t which has no direct experimental counterpart.

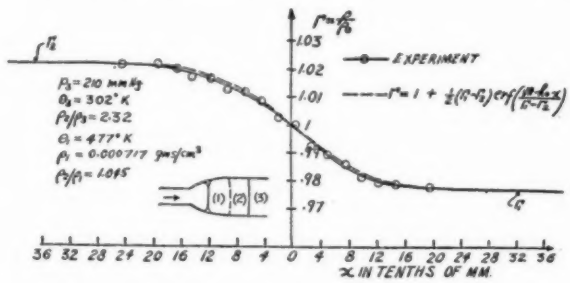


FIG. 5. Typical results obtained for density transition through weak type B contact region.

Nevertheless, a qualitative comparison was made using the theoretical t as that defined by (10) with $s(t)$ taken as the experimental thickness. If $h_0(t)$ is the maximum negative slope of the density profile, then for weak transitions

$$\Gamma=1+\tfrac{1}{2}(\Gamma_1-\Gamma_2)\,\mathrm{erf}\left(\frac{\pi^{\frac{1}{3}}h_0}{\Gamma_1-\Gamma_2}x\right) \tag{14}$$

from (9) and (10).

Figure 5 shows a typical plot obtained of experimental density change and that computed from (14) with h_0 taken as the maximum experimental slope. The qualitative agreement suggests that (14) may be a reasonable approximation for weak transitions.

ACKNOWLEDGMENTS

The author is indebted to Professor G. N. Patterson, Institute of Aerophysics, University of Toronto, for suggesting this problem. Financial support was received from the Defence Research Board of Canada, and the Wallberg Research Fellowship, University of Toronto.

Short-Time Stress Relaxation Behavior of Plastics*

M. T. WATSON, W. D. KENNEDY, AND G. M. ARMSTRONG
Research Laboratories, Tennessee Eastman Company, Division of Eastman Kodak Company, Kingsport, Tennessee
(Received September 23, 1954)

An apparatus for measuring stress relaxation of plastics at constant strain in the time range of 0.01 to 2.5 sec after load application is described. Experimental results obtained indicate that rigid plastics at room temperature may be classified as either brittle or ductile. Brittle plastics sustain only low strains without fracture at the testing speed of the apparatus and undergo little relaxation of stress in this time range; ductile plastics sustain higher strains and undergo considerable relaxation of stress. The factors affecting relaxation behavior are briefly discussed. Increasing the strain, temperature, or plasticizer content generally increases the relaxation rate.

Many of the results yield a linear plot of stress *vs* logarithmic time, as has been reported in the literature for various materials tested at longer times. By application of one form of the Eyring absolute rate theory, an average free energy of activation for the relaxation process can be calculated. The value thus calculated for polymethyl methacrylate is in the range reported in the literature for other polymers.

An empirical measure of "toughness" can be based on these results. For many plastics this measurement is in accord with service performance.

INTRODUCTION

THE measurement of stress relaxation at constant strain is becoming an increasingly important method of studying polymer behavior. Most of the published measurements are ones in which several seconds, or even minutes, elapsed before the first measurement of stress was made.[1] In some cases, testing times have been extended to many hours.[2] For practical application, as well as for fundamental study, relaxation behavior of plastics at short times is also of considerable interest.

In this paper an apparatus for following the short-time stress relaxation behavior of plastics is described. Experimental results obtained with several commercial plastics, including both brittle and ductile types, are presented. The effects of strain, temperature, plasticizer content, and specimen shape on the relaxation behavior were studied. Most of the results yield a linear plot of stress *vs* logarithmic time. Such a relationship is predicted by the Eyring absolute rate theory, from which an average free energy of activation for relaxation is calculated from our results on polymethyl methacrylate. An empirical "toughness" measure based on these relaxation results is discussed briefly.

APPARATUS

Figure 1 shows a schematic diagram of the apparatus. The essential mechanical parts comprise a mounting frame sufficiently rigid that its deformations under stress are negligible compared with those of the specimen; clamps for containing the specimen; a massive spring whose release effects sudden elongation of the specimen; a spring bar which transmits the force of the spring to the clamp and specimen; a means of suddenly releasing the spring; a set of blocks which stops the spring motion and, by proper selection, permits extension of the specimen to a predetermined amount; a hydraulic jack for cocking the spring; and, not shown in the figure, a thermostatic enclosure around the specimen, permitting measurements at various temperatures.

The massive spring is confined between a fixed part of the apparatus and the enlarged lower end of the spring bar. Motion of the jack is transmitted through the spring bar extension to the spring bar to compress the spring, which is then cocked in the compressed position by rotating the collar so that its three hardened pins engage in a circumferential slot around the spring bar. Release of the spring to elongate the specimen can then be effected by rapid rotation of the collar by means of compressed air to align the pins with three longitudinal slots in the spring bar intersecting the circumferential slot. Fixed keys in the longitudinal slots prevent twisting of the spring bar. Specimen extension up to as much as $\frac{1}{8}$ in. or more is thus achieved in times of the order of 2 or 3 msec, exclusive of the time taken for travel of the pins into alignment with the longitudinal slots.

The load-measuring device is a tubular aluminum tensile weighbar mounting four electric resistance-wire strain gauges wired in a bridge network; specimen extension produces a dc unbalance of the bridge proportional to the load. This load signal is then suitably amplified, fed simultaneously to one channel of a Brush Model BL-202 double-channel magnetic oscillograph, and to a DuMont 208B cathode-ray oscilloscope through its direct-coupled amplification stage only.

At present, our most sensitive weighbar is a tube of

* Presented in part at the American Physical Society meeting in New York, February 1–3, 1951, and at the XIIth International Congress of Pure and Applied Chemistry in New York, September 10–13, 1951.

[1] H. D. Smith and R. Eisenschitz, J. Textile Inst. **22**, T170 (1931); B. Maxwell and L. F. Rahm, Ind. Eng. Chem. **41**, 1988 (1949); S. M. Katz and A. V. Tobolsky, Textile Research J. **20**, 87 (1950); E. Catsiff and A. V. Tobolsky, J. Appl. Phys. **25**, 1092 (1954).

[2] Tobolsky, Prettyman, and Dillon, J. Appl. Phys. **15**, 380 (1944); E. E. Wright, ASTM Bull. **184**, 47 (1952); Chen, Ree, and Eyring, Textile Research J. **22**, 416 (1952).

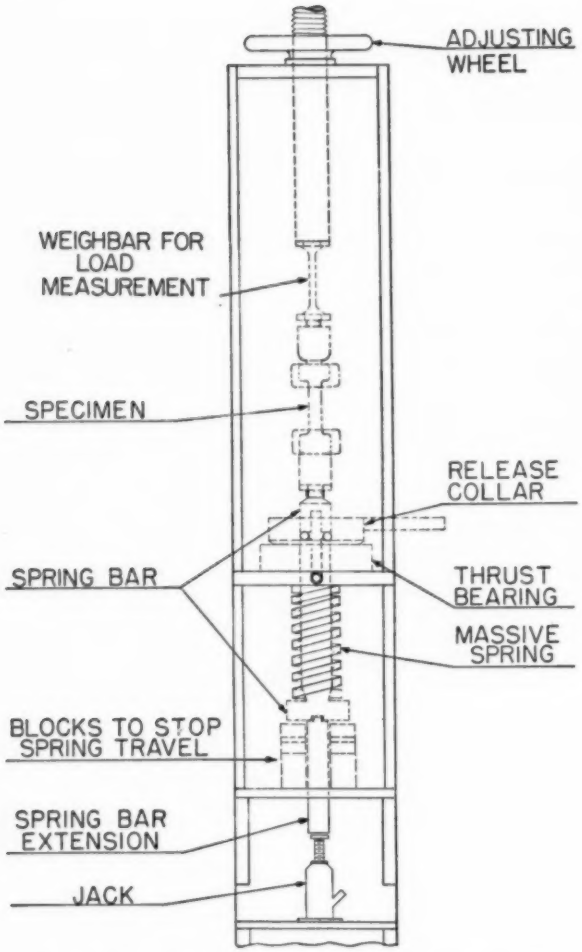


FIG. 1. Schematic diagram of stress relaxation apparatus.

Alcoa 14S-T6 aluminum, 0.484 in. i.d. and 0.534 in. o.d. Baldwin SR-4 type C-14 strain gauges are used; each has a resistance, R, of 2000 ohms and a gauge factor, k, of about 3.3. The sensitivity coefficient, S, of the weighbar, which relates load to measured bridge unbalance by the relation $P=S(\Delta V/i)$, is defined as $S=2AE/kR[1+\mu]$. The calculated sensitivity, S, is 98.5 lb-ma per mv for this particular weighbar, using $E=1.08\times10^7$ psi and μ (Poisson's ratio) $=0.33$. An experimental value of 86.9 was obtained from calibration measurements.

Figure 2 shows a typical cathode-ray oscillogram for a plasticized cellulose acetate butyrate and a typical Brush oscillograph record for a similar material. The load-time curve of the latter has been inked-in for clarity. The frequency of the sinusoidal trace in Fig. 2(a) is 200 cps; thus, excellent resolution is obtained within the first 100 msec after elongation. The Brush oscillograph records enable the load decrease to be followed for several seconds [see Fig. 2(b)]. Loads in the range of 10 to 1000 lb are measurable with an accuracy of about 2 percent.

Strain was measured with an extensometer fabricated by bending a $\frac{5}{8}$-in. wide strip of 0.020-in. thick SAE 1070 or 1085 spring steel into the shape of a bicycle clip of about $\frac{3}{4}$-in. radius. Four Baldwin SR-4 type C-14 strain gauges were cemented to the extensometer,

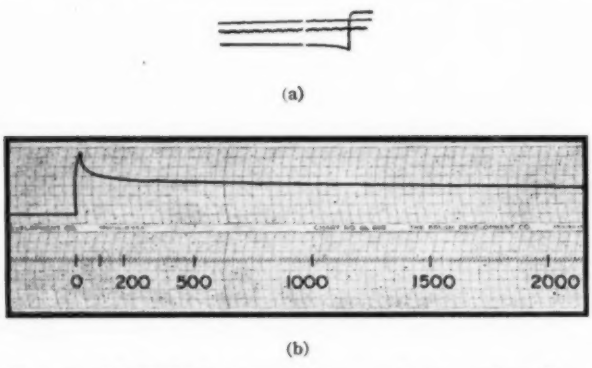


FIG. 2. (a) Cathode-ray oscillogram of stress relaxation of plasticized cellulose acetate butyrate. (b) Magnetic oscillogram of stress relaxation of plasticized cellulose acetate butyrate. Time units are milliseconds.

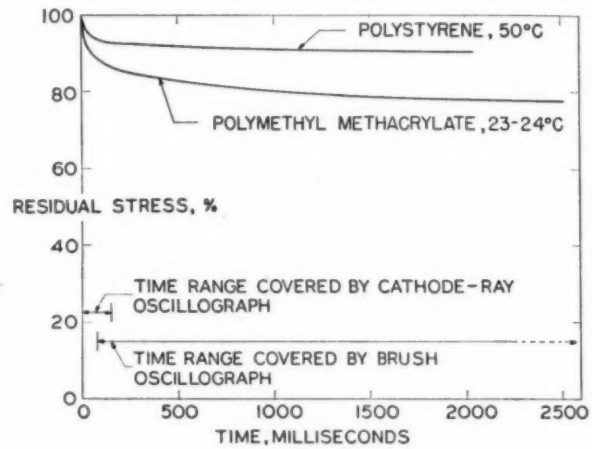


FIG. 3. Stress relaxation rates of brittle polymers at 0.7 percent strain.

which was then clamped to the specimen to follow its elongation. The strain signal was recorded as a dc voltage on a second channel of the Brush magnetic oscillograph. This extensometer had a calculated sensitivity of 50 mv per ma per in. deflection; calibration gave an experimental value of 62.0.

Measurements of stress relaxation in the approximate time range of 0.01 to 2.5 sec were thus made on a number of plastics under different testing conditions.

EXPERIMENTAL RESULTS

Brittle Plastics

Materials which are hard, stiff, and brittle were found to be characterized by little relaxation within the time range of measurement; the strains which these materials sustained without fracture were small. Thus, commercial polystyrene at room temperature did not withstand strains as great as 1 percent at the testing

speed of the apparatus without fracture. The relaxation behavior of two such brittle plastics is shown in Fig. 3: a commercial polystyrene at 50°C and a commercial polymethyl methacrylate at room temperature. Strain was approximately 0.7 percent. The data are plotted here as percentage of residual stress *vs* time where residual stress = (instantaneous stress/peak stress) ×100.

Ductile Plastics

Softer, more ductile plastics, on the other hand, have faster relaxation rates and are considerably more extensible than the brittle plastics mentioned. In our apparatus, materials like nylon, polyethylene, and plasticized cellulose esters, for example, sustained strains of at least 5–10 percent without fracture and underwent marked relaxation of stress within a few

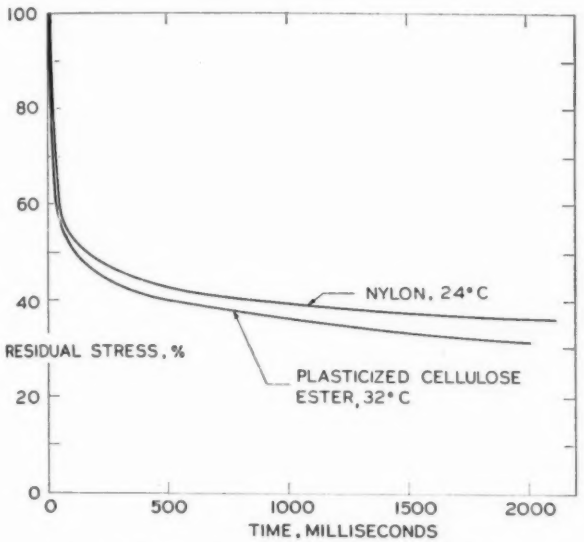


FIG. 4. Stress relaxation rates of ductile polymers at 5.5–6.0 percent strain.

seconds. Figure 4 shows relaxation curves for two typical ductile plastics: a commercial nylon and a plasticized cellulose acetate butyrate at 5.5–6 percent strain and at temperatures of 24°C and 32°C, respectively. Stress decay is seen to be rapid.

Factors Affecting Relaxation

The relaxation behavior of a plastic is a function not only of the composition of the material but also of the strain. The effect of varying the strain depends upon the nature of the plastic. Thus, for plasticized cellulose acetate butyrate, relaxation was markedly increased by an increase in strain at 30–32°C, as shown in Fig. 5. For polyethylene, on the other hand, the relaxation rate was nearly independent of strain over the range of 2 percent to 8 percent strain.

In general, the effect of an increase in temperature is to increase the relaxation rate, but the magnitude of

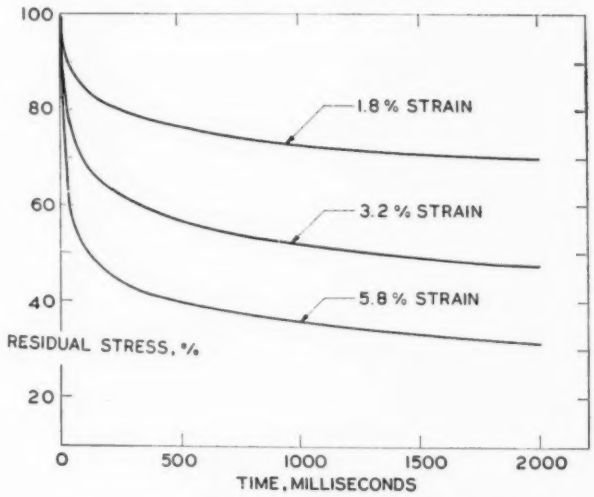


FIG. 5. Stress relaxation rates of plasticized cellulose acetate butyrate at 30–32°C.

the temperature dependence differs for various plastics. This effect is particularly pronounced with plasticized cellulose acetate butyrate, for which relaxation curves are shown in Fig. 6, at about 3.2 percent strain and temperatures of 0° and 32°C. At lower strains, the influence of temperature is considerably less.

Both the kind and the amount of plasticizer influence the relaxation behavior of plastics. The relaxation rate increases with increasing plasticizer content. The magnitude of the effect depends on the temperature and on the strain.

Another factor which affects the relaxation behavior is the shape of the specimen. Most of our measurements were made with injection-molded specimens having a dumbbell-shaped profile, with a 1-in. length of $\frac{1}{4}$-in. square cross section flaring at the ends into a $\frac{3}{4}$-in. square section which was held in the clamps. However, there was a noticeable difference in relaxation behavior among four specimen types, all prepared from the same batch of commercial polymethyl methacrylate. Besides

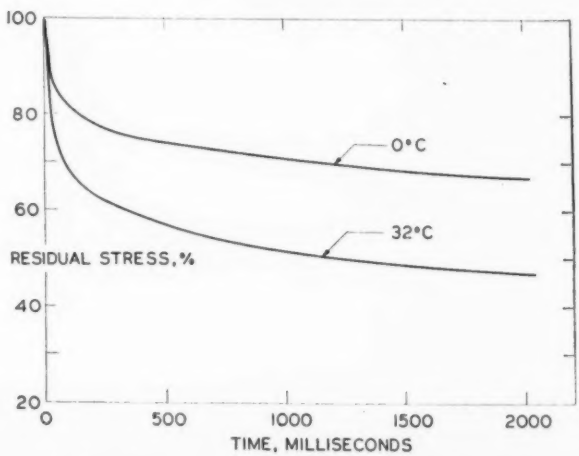


FIG. 6. Stress relaxation rates of plasticized cellulose acetate butyrate at 3.2 percent strain.

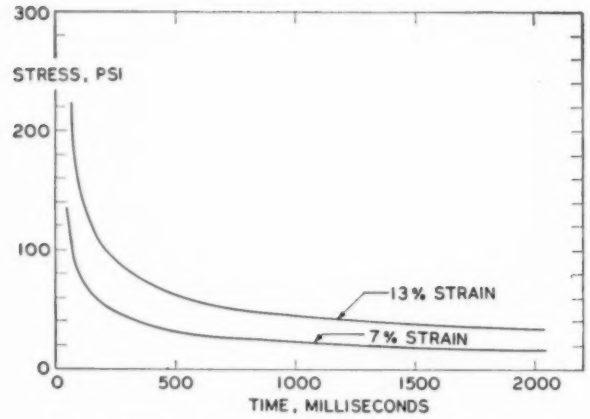


FIG. 7. Stress relaxation of polyisobutylene at −40°C.

the standard specimen, the three other types were (1) injection-molded, but with a $\frac{1}{8}$- by $\frac{1}{4}$-in. cross section; (2) injection-molded with a 1-in. gauge length of circular cross section of $\frac{1}{4}$-in. diameter; and (3) standard-size square cross section, milled from a $\frac{1}{4}$-in. compression-molded sheet.

Polyisobutylene

We have participated in a cooperative program on dynamic testing methods sponsored by the National Bureau of Standards, under the direction of Dr. Robert S. Marvin. The standard sample sent to the participating laboratories was a polyisobutylene having a viscosity-average molecular weight of 1.35 million.[3] For such a soft, rubbery material, it was necessary to use a testing temperature of −40°C in order to attain a measurable stress level. A plot of stress *vs* time for this material at this temperature and at strains of about 7 percent and 13 percent is shown in Fig. 7. The material is seen to have a very rapid short-time decay of stress. Because of the high gain of the amplifiers necessary for such a low stress level, results at shortest times were obliterated by high-frequency mechanical noise.

DISCUSSION

It is difficult to make direct comparisons of our results with those of other workers because few relaxation measurements on rigid plastics have appeared in the literature. A recent article by Tobolsky and co-workers,[4] however, gives results of measurements on polyethylene and permits a comparison of our results with theirs and with dynamic measurements of Dunell and Dillon[5] on this material. According to Tobolsky,[4] if stress relaxation results yield a modulus *vs* log time plot which is linear over a "wide range of" logarithmic time, then a value of $\omega\eta$ can be calculated from the slope of this plot and compared with values of the same quantity measured in dynamic experiments. We have evaluated this quantity $\omega\eta$ from our relaxation measurements on polyethylene, plotted as stress *vs* logarithmic time in Fig. 8. Table I shows a comparison of our results with those of Tobolsky and with experimental values determined by Dunell and Dillon in forced longitudinal vibration experiments. Our values agree reasonably well with those from stress relaxation measurements of Tobolsky, but both sets of values are lower than those obtained in dynamic measurements.

Our results on polymethyl methacrylate, shown in Fig. 3, can be compared with results of relaxation experiments on this material recently published by McLoughlin and Tobolsky.[6] Figure 9 is a plot of modulus *vs* logarithmic time of our results, calculated using an average strain value of 0.7 percent and those of McLoughlin and Tobolsky. Although both sets of data give linear plots, the two curves do not coincide;

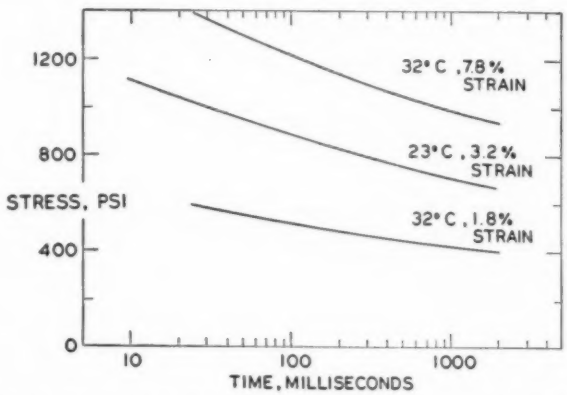


FIG. 8. Stress relaxation, semilogarithmic plots, of polyethylene.

in fact, the slopes are slightly different. Our measurements were made with no special preconditioning treatment. McLoughlin and Tobolsky[6] pointed out, however, that reproducible pretreatement is necessary since water seems to plasticize polymethyl methacrylate, at least under certain conditions. Furthermore, the rate of cooling affects the relaxation behavior.[6,7] Our results were obtained on injection-molded samples which were not reproducibly cooled. It seems likely that these two effects are responsible for the discrepancy shown in Fig. 9.

Our data do not yield a simple exponential stress-time function. In many cases, the stress is a linear function (within 5 percent of less) of the logarithm of the time over the $2\frac{1}{2}$ or so cycles of logarithmic time that the measurements cover, as illustrated in Figs. 8 and 9. This is true for a variety of plastics tested, including both brittle and ductile materials.

[3] R. S. Marvin, "Interim report on the cooperative program on dynamic testing," Natl. Bur. Standards (1951).

[4] Tobolsky, Dunell, and Andrews, Textile Research J. **21**, 404 (1951).

[5] B. A. Dunell and J. H. Dillon, Textile Research J. **21**, 393 (1951).

[6] J. R. McLoughlin and A. V. Tobolsky, J. Colloid Sci. **7**, 555 (1952).

[7] J. R. McLoughlin and A. V. Tobolsky, J. Polymer Sci. **7**, 658 (1951).

The linear dependence of stress on the logarithm of the time has been observed, at considerably greater values of the time, for many materials. In an early paper, Phillips[8] reported such a relationship, over times of the order of a day, for India rubber in tension. Macdonald and Ushakoff[9] obtained linear stress-log time curves for a variety of substances, in tension and in compression. Finch[10] reported similar data from compression tests on cotton, wool, and viscose fibers.

Such a relationship between stress and time also can be deduced from the absolute reaction rate theory of flow as applied by Eyring and co-workers to a simple 3-element mechanical model of a relaxing polymer.[11] For small strains and for times less than 1000 min, the following approximate relation can be deduced[12]:

$$\sigma=\sigma_0-A\log(1+Bt) \qquad (1)$$

where σ is the stress at any time t, σ_0 is the initial zero-time stress, and A and B involve various parameters of the relaxing system. Although the 3-element model includes no provision for permanent set and the and the treatment does not include a distribution of

TABLE I. $\omega\eta_{dyn}$. values for polyethylene, dynes/cm².

Calculated from stress relaxation data of Tobolsky, *et al.*[a]	Calculated from our stress relaxation data	Experimental value, dynamic measurements of Dunell and Dillon[b]
$\epsilon^{(a)}=5\%$, 0.18×10^9 $\epsilon=0.5\%$, 0.20×10^9	$\epsilon=7.8\%$, 0.14×10^9 $\epsilon=3.2\%$, 0.28×10^9 $\epsilon=1.8\%$, 0.26×10^9	0.82×10^9

(a) ϵ=strain

[a] See reference 4.
[b] See reference 5.

elements, Eq. (1) can be applied to experimental results to obtain an average free energy of activation for the relaxation process. By making certain assumptions, we obtained from the slope of the stress-log time plot of our data on polymethyl methacrylate (Fig. 9) a value of about 27 kg-cal. This value is in reasonable

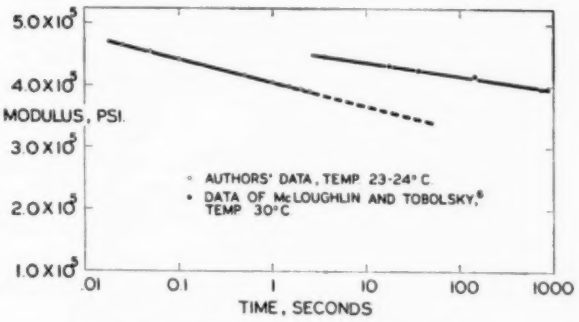


FIG. 9. Stress relaxation, semilogarithmic plots, of polymethyl methacrylate.

agreement with the value of 23 kg-cal calculated by Lyons[13] as an average activation energy for viscoelastic deformation, employing a "box" distribution of relaxation times.[4] Our value is also in the range of activation energy values for various other polymers reported in the literature.[11,14]

An empirical measure of "toughness" can be based on these relaxation data. Under a given set of test conditions, the tougher the material the faster is its relaxation rate. For many plastics, a correlation exists between impact strength and some measure of the speed of stress decay, for example, the percentage of relaxation that has taken place after some arbitrary time. In a few instances, impact data fail to rate plastics in the proper order of toughness based on known service performance, but the anomaly is resolved by comparison of relative relaxation rates. It should be emphasized, however, that it is an oversimplification to consider relative relaxation rates alone as a measure of toughness. Any numerical factor by which plastics could be rated as to toughness would have to include other quantities, for example, the maximum deformation sustainable without fracture and the stress developed in producing a given strain.

ACKNOWLEDGMENT

The authors would like to acknowledge the important contributions of Mr. T. A. Mitchell, who did most of the design work on the apparatus.

[8] P. Phillips, Proc. Phys. Soc. (London), **19**, 491 (1905).
[9] W. B. Macdonald and A. Ushakoff, J. Colloid Sci. **3**, 135 (1948).
[10] R. B. Finch, Textile Research J. **18**, 237 (1948).
[11] Halsey, White, Jr., and Eyring, Textile Research J. **15**, 295 (1945).
[12] S. Yurenka (private communication).
[13] W. J. Lyons, J. Appl. Phys. **24**, 217 (1953).
[14] Eyring, Alder, Rossmassler, and Christensen, Textile Research J. **22**, 223 (1952); A. V. Tobolsky and H. Eyring, J. Chem. Phys. **11**, 125 (1943).

Microwave Diffraction by Apertures of Various Shapes*

S. J. Buchsbaum,† A. R. Milne,‡ D. C. Hogg,§ G. Bekefi, and G. A. Woonton
Eaton Electronics Research Laboratory, McGill University, Montreal, Canada

(Received December 15, 1954)

Experimental results of the near zone diffraction field of circular and elliptical apertures and of the infinite slit are presented. The investigations were made on apertures which were illuminated by a plane polarized electromagnetic wave and whose characteristic dimensions were varied over a considerable range with respect to the wavelength of the incident radiation.

Most of the measurements are compared with predictions from recently published approximate theories and in a few cases with exact solutions. In general the agreement between theory and experiment is found to be satisfactory. Some of the discrepancies are discussed.

I. INTRODUCTION

IT is known that the classical scalar Kirchhoff-Fresnel[1] diffraction theory and the more recent Stratton and Chu[2] vector representation fail to predict the field within and in the immediate vicinity of apertures cut in large plane perfectly conducting screens. This has led several writers[3–11] within the last decade to propose approximate solutions to the problem with the view to predicting adequately the near as well as the far field distributions. During the Symposium on Microwave Optics held at McGill University in June 1953, some of these approximate theories were discussed; it was pointed out among other things that they had not been tested adequately, because until that time near field measurements on circular apertures only (and on the complementary disk) had been made.[12–15] It was suggested further that the reported good agreement between theory and experiment in the case of the circular aperture perhaps was the result of the special symmetry of the aperture and not to the excellence of the theory.

It is not necessary to stress here the need for an approximate theory describing with sufficient accuracy the diffraction field of an aperture of arbitrary shape and size; the engineer and the applied physicist have a definite need for such a theory. A general exact solution to the problem does not exist except in an abstract form.[16] The few exact solutions now in existence for apertures of special shapes employ complicated functions of mathematical physics (e.g., oblate spheroidal functions in the case of the circular aperture[17] and Mathieu functions for the infinite slit[18]) which are little known and which are tabulated over a very limited range of argument and order at best. Calculations are made even more difficult by the fact that the solutions appear as slowly convergent series with the result that the method fails when the characteristic aperture dimensions exceed one or two wavelengths of the incident radiation.

At the time of the Microwave Symposium the work of measuring the aperture and near diffraction fields of long slits and ellipses of various shapes was already in progress. A comparison of the results obtained with calculations based on approximate and exact theories was the next natural step. In addition to measurements made on slits and elliptical apertures, results are presented for circular holes varying in size from one-quarter to forty wavelengths in diameter; these latter investigations supplement and extend those reported by other workers.[12–15]

Of the several approximate theories, only one will be used for comparison with experiment. It was derived independently by Severin,[3] Silver and Ehrlich,[7] Neuge-

* This work was supported at McGill University by the United States Air Force through its Cambridge Research Center under Contract AF 19(122)–81.

† Present address: Department of Physics, Massachusetts Institute of Technology, Cambridge, Mass.

‡ Present address: Pacific Naval Laboratory, Esquimalt, B. C., Canada.

§ Present address: Bell Telephone Laboratories, Holmdel, New Jersey.

[1] See for instance: B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens' Principle* (Oxford at the Clarendon Press, 1950), second edition, Chap. II.

[2] J. A. Stratton and L. J. Chu, Phys. Rev. **56**, 99 (1939).

[3] H. Severin, Z. Naturforsch. **1**, 487 (1946); also presented at the Symposium on Microwave Optics, McGill University, June, 1953.

[4] W. Braunbek, Z. Physik **127**, 381 (1950); **127**, 405 (1950).

[5] W. Franz, Z. Physik **128**, 432 (1950); Proc. Phys. Soc. (London) **A63**, 925 (1950).

[6] W. Frahn, thesis, University of Aachen (1951).

[7] S. Silver and M. J. Ehrlich, Antenna Laboratory Report No. 181 (1951), Department of Engineering, University of California.

[8] H. E. J. Neugebauer, Eaton Electronics Research Laboratory Report No. b5 on Contract AF 19(122)–81 to the U. S. Air Force, Cambridge Research Center (1952); also J. Appl. Phys. **23**, 1406 (1952).

[9] G. Bekefi, J. Appl. Phys. **24**, 1123 (1953); **23**, 1403 (1952).

[10] G. A. Wooton, J. Appl. Phys. **23**, 1405 (1952).

[11] For an extensive review of diffraction theories see C. J. Bouwkamp, Repts. Progr. in Phys. **17**, 35 (1954).

[12] C. L. Andrews, Phys. Rev. **71**, 777 (1947); J. Appl. Phys. **21**, 761, (1950).

[13] H. Severin, Z. Naturforsch. **1**, 487 (1946).

[14] S. Silver and M. J. Ehrlich, Antenna Laboratory Report No. 185 (1952), Department of Engineering, University of California.

[15] G. Bekefi and G. A. Woonton, The Eaton Electronics Research Laboratory Report No. b7 under contract to the U. S. Air Force, Cambridge Research Center.

[16] E. T. Copson, Proc. Roy. Soc. (London) **A186**, 100 (1946); also N. Marcuvitz, Research Report No. EM–50, Washington Square College of Arts and Science, Mathematics Research Group, New York University, Chapter X.

[17] J. Meixner and W. Andrejewski, Ann. Physik (6) **7**, 157 (1950); also W. Andrejewski, Naturwiss. **38**, 406 (1951); Z. angew. Phys. (5) **5**, 178 (1953).

[18] P. M. Morse and P. J. Rubinstein, Phys. Rev. **54**, 895 (1938).

bauer,[8] and Bekefi[9]; the basic simplifying assumptions made by these authors in deriving their respective formulas are essentially the same. The final mathematical expressions, however, restricted in many cases to special conditions of the illuminating wave and of the aperture shape, differ in form and in the degree of their mathematical complexity. It has been shown elsewhere[9] that the theory can be formulated as follows: consider an infinitely large plane, perfectly conducting screen of negligible thickness situated in the x, y plane of a rectangular coordinate system (Fig. 1), pierced by a hole of arbitrary shape whose area is designated by S_0. A monochromatic wave polarized in the x direction, originating from a source situated in the negative half space ($z<0$), is assumed to be incident upon the screen. The diffracted field at any point (x,y,z) in the positive half space ($z\geq 0$) is then described by that component of the Hertz vector $\mathbf{\Pi}$ which lies in the direction of the initial polarization and is given by

$$\Pi_x(x,y,z)=-\frac{1}{2\pi}\int_{S_0}\left[\frac{\partial\Pi_{ix}(x',y')}{\partial z'}\right]_{z=0}\times\frac{\exp(-jkr)}{r}dx'dy' \quad (1)$$

with $\Pi_y=\Pi_z=0$. The subscript "i" denotes the incident field that would be present were the screen removed, $(\partial\Pi_{ix}/\partial z)$ is the normal derivative of Π_{ix} and the integration extends over the area of the aperture S_0 only; r represents the distance of the field point (x,y,z) from an element $(\delta x'\delta y')$ of the aperture and k is the wave number $2\pi/\lambda$ where λ is the wavelength. Harmonic time variation $\exp(j\omega t)$ is assumed implicitly, where ω is the angular frequency.

The electric and magnetic field components $\mathbf{E}$ and $\mathbf{H}$ can then be obtained from the well-known relations

$$\mathbf{E}=k^2\mathbf{\Pi}+\nabla\nabla\cdot\mathbf{\Pi} \quad (2a)$$

$$\mathbf{H}=-\frac{k^2}{j\omega\mu}\nabla\times\mathbf{\Pi}, \quad (2b)$$

expressed in the mks system of units.

In the derivation of Eq. (1) the assumption was made that the component of the magnetic field tangential to the screen can be set equal to zero on its shadow side. Experiment indicates that this postulate is approximately true (except at points very close to the diffracting edge), provided that the aperture is of dimension sufficiently large as compared to the wavelength. The lower limit of aperture size for which this approximation is valid can be determined by field measurements alone (for the circular aperture the smallest hole is about one wavelength in diameter).

It should be pointed out further that although Eq. (1) satisfies (a) Maxwell's field relations everywhere in free space, and (b) the conditions of continuity of the field components across the diffracting aperture (as opposed to Kirchoff and Stratton and Chu's solutions which fail in this respect), it does not obey the edge conditions of Meixner[19] and Bouwkamp[20] and in this aspect it suffers from the same defect as the classical formulas.

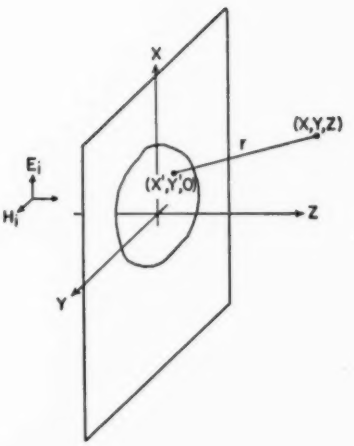


FIG. 1.

Various degrees of difficulty are encountered in the evaluation of the integral appearing in Eq. (1), depending on the shape and size of the aperture and on the position of the field point. For the actual methods of calculation employed, the reader is referred to the original reports and papers.[8,9,15,21,22]

II. EXPERIMENTAL RESULTS AND COMPARISON WITH THEORY

Most of the measurements were made at microwave frequencies corresponding to 3.20 and 1.25 cm wavelengths. At these frequencies only the electric component of the electromagnetic wave was examined; the probe was a balanced half-wave dipole whose characteristic receiving pattern had been determined previously. It is shown elsewhere[23] that a field probe of this type does not appreciably disturb the field to be measured and that the results are trustworthy to better than 0.5 decibel except at points lying very close to the diffracting edge. The apertures which were cut in a large, thin, metal screen were illuminated by plane polarized electromagnetic waves issuing from a rectangular horn fed by a klystron generator.

The experiments on the very small circular holes were made at a wavelength of 50 cm; in consequence, it was possible to use extremely small probes and to

[19] J. Meixner, Ann. Physik **6**, 1 (1949).
[20] C. J. Bouwkamp, Physica **12**, 467 (1946).
[21] A. R. Milne, M.Sc. thesis, McGill University, Montreal, Canada (1953).
[22] S. J. Buchsbaum, Eaton Electronics Research Laboratory Report No. b9 on contract to the U. S. Air Force, Cambridge Research Center (1953).
[23] G. A. Woonton, Eaton Electronics Research Laboratory Report No. a2b4 on contract to the U. S. Air Force, Cambridge Research Center (1952); also R. B. Borts and G. A. Woonton, Eaton Electronics Research Laboratory Report No. a3 on contract to the U. S. Air Force, Cambridge Research Center (1952).

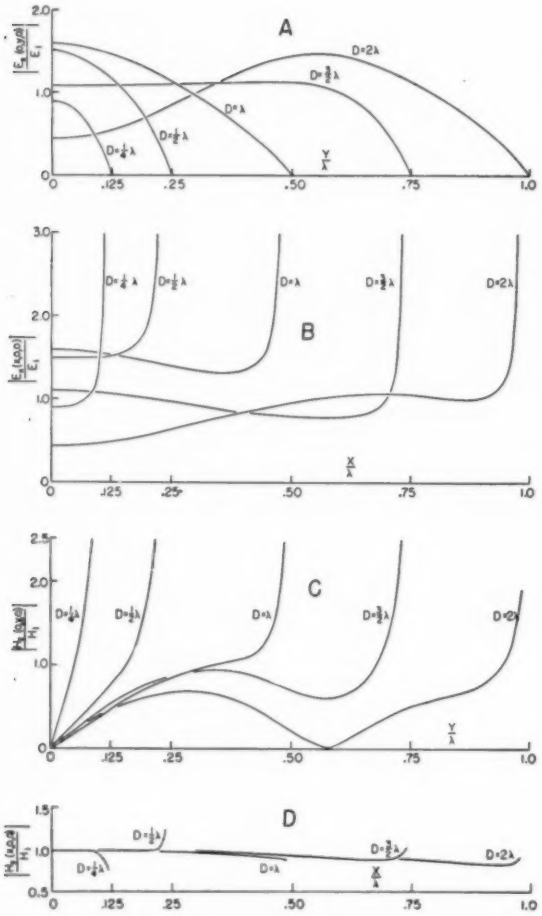


FIG. 2. Measured amplitude of the electric and magnetic field distributions within circular apertures of diameters D from 0.25λ to 2.0λ.

make measurements very close to the diffracting rim. The balanced electric dipole used to measure the electric component and the shielded loop employed for magnetic measurements were of dimensions less than one-thirtieth of a wavelength. The apertures were irradiated by plane polarized waves from a paraboloidal reflector which was 60 inches in diameter.

In all cases measurements were performed in planes parallel to that of the screen (i.e. the x, y plane of Fig. 1) at various distances z from the screen with z ranging from $z=0$ (aperture field measurements) to $z=10\lambda$. The field distributions were examined in the two mutually perpendicular planes known as the E and H planes; in the former the scan was made in the direction of the initial incident polarization, while in the latter the scan was perpendicular to it.

The field distributions were recorded automatically. A motor-driven optical bench carried the probe through the field. The optical bench was built from a six-foot lathe bed and the probe carriage was driven along the ways of the lathe by a lead screw. Rotation of the lead screw was transmitted through appropriate gearing to a Selsyn generator; by this means the motion of the paper through a type 373 Airborne Instruments Laboratory rectangular coordinate recorder was made synchronous with the motion of the probe traversing the field.

The audio modulation signal of the rectified output was amplified by a narrow-band amplifier. The amplified audio component served to control the pen positioning circuits of the A.I.L. recorder. The combined motion of pen and paper resulted in a graph of the relative power at the probe terminals, plotted on a decibel scale, as function of the position of the probe.

Free-space measurements were made with the screen removed, in order to determine the planeness of the incident wave and also to enable the intensity of the diffraction field to be normalized to that of the incident wave. It was found that over an area at least as great as that occupied by the largest aperture that was used, the power was effectively constant and did not fluctuate by more than ±0.3 decibel about the mean level.

With the electric component of the incident wave taken along the x-axis of the coordinate system, the

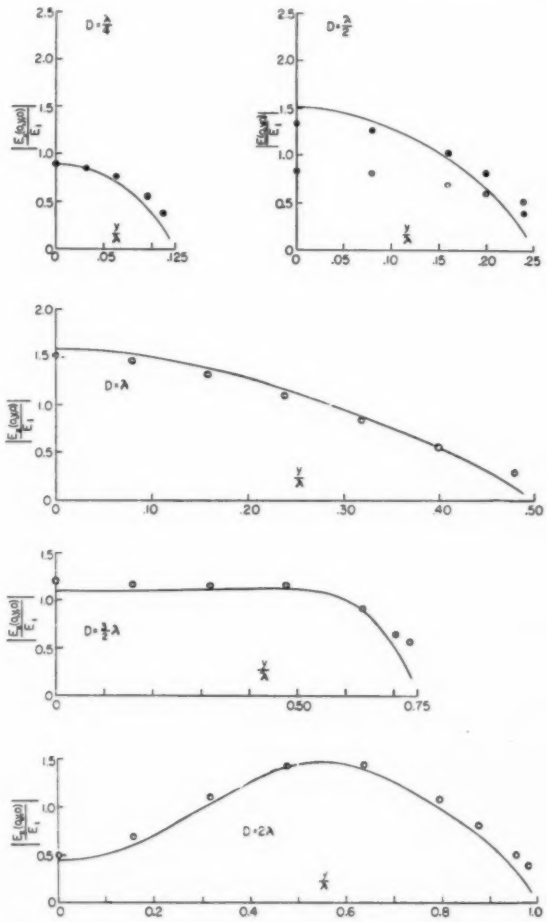


FIG. 3. Amplitude of the electric field within circular apertures $D=0.25\lambda$ to $D=2.0\lambda$. H-plane distribution. [—— measured; ●● calculated from Bouwkamp's small hole theory; ○○ calculated from Eq. (1).]

nomenclature describing the ordinates of the various graphs presented, will become obvious on reference to Fig. 1.

(a) The Circular Aperture

In Fig. 2 are compiled aperture field measurements ($z=0$) performed at a wavelength of 50 cm on circular apertures of diameters D from 0.25λ to 2λ (i.e., from $ka=0.25\pi$ to 2π where a is the radius); all patterns were normalized to the intensity of the incident, plane unperturbed wave (denoted by subscript "i"). The graphs designated by A and B are the E and H plane patterns of the electric vector while those denoted by C and D are the normal and tangential components of the magnetic vector, respectively. It should be pointed out that from conditions of continuity of the field components through the aperture, the theoretical requirement follows that for a thin perfectly conducting screen, the tangential component of the magnetic vector (H_y) within the aperture is identically equal to that of the unperturbed incident wave that would exist there

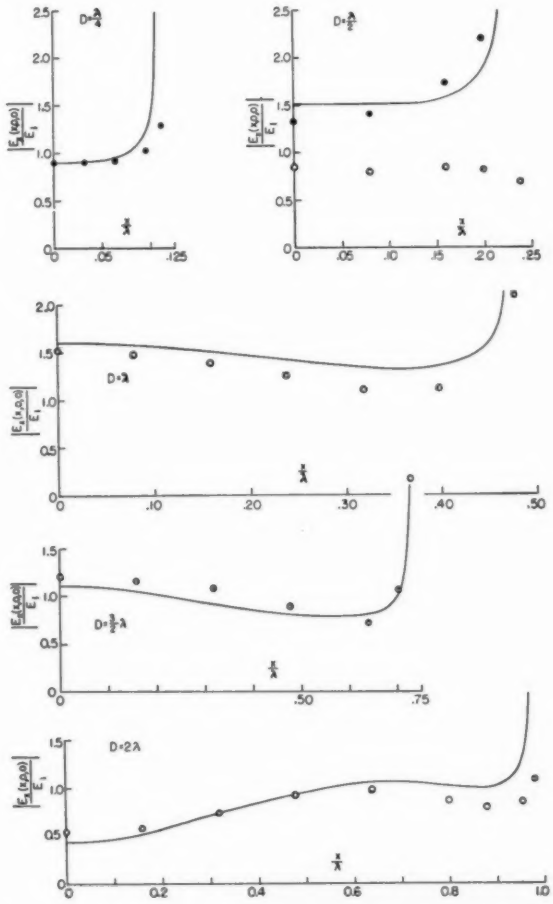


FIG. 4. The electric field within circular apertures, $D=0.25\lambda$ to $D=2.0\lambda$. E-plane distribution. [—— measured; ●● calculated from Bouwkamp's small hole theory; ○○ calculated from Eq. (1).]

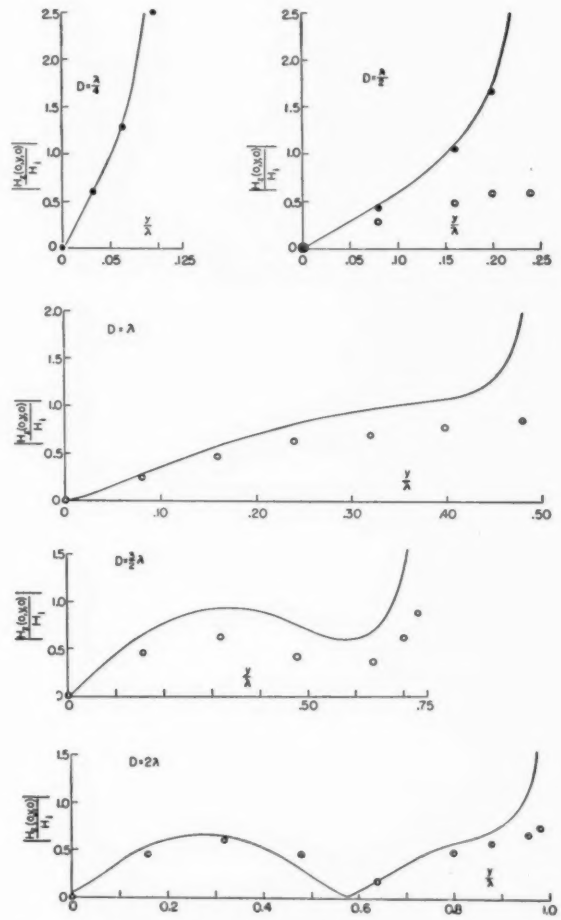


FIG. 5. Amplitude of the normal component of the magnetic field within circular apertures, $D=0.25\lambda$ to $D=2.0\lambda$. H-plane distribution. [—— measured; ●● calculated from Bouwkamp's small hole theory; ○○ calculated from Eq. (1).]

in the absence of the screen.[14,24] That this is nearly so experimentally can be seen from Fig. 2(D). It will be noted that some of the curves in this figure have negative while others have positive curvature at the very edge of the aperture. These effects are believed to be due to the fact that in the extreme proximity of the diffracting rim, where a strong component of electric field exists [Fig. 2(B)], the magnetic probe, being somewhat sensitive to electric field could pick up this signal and yield erroneous results, the nature of which would depend critically upon the exact position and orientation of the gap in the shield of the probe.

In Figs. 3–5 are shown some of the measurements of Fig. 2 compared with theoretical predictions. Figure 3 gives H-plane patterns of the electric vector for the various apertures, Fig. 4 the corresponding E plane patterns, and Fig. 5 the distribution of the normal component of the magnetic vector of the electro-

[24] S. Silver, Antenna Laboratory Report No. 163 (1949), Department of Engineering, University of California. This was deduced originally by H. A. Bethe, Phys. Rev. **66**, 163 (1944) and E. T. Copson (reference 16).

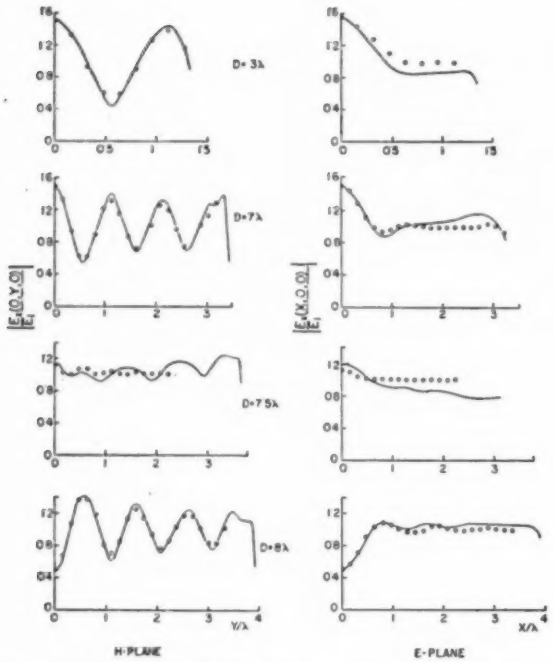


FIG. 6. The electric field within circular apertures, $D=3\lambda$ to $D=8\lambda$. E- and H-plane distributions. [—— measured; ○○ calculated from Eq. (1).]

magnetic field, all measured within the aperture ($z=0$). For the two smallest apertures ($D=0.25\lambda$ and $D=0.5\lambda$) comparisons are made with the small hole theory of Bouwkamp[25] (solid circles). It is seen that the agreement with theory is good even though the latter is not strictly applicable to the larger of the two apertures ($D=0.5\lambda$). The agreement with predictions obtained from Eq. (1) (open circles) is very satisfactory except when the aperture diameter is less than one wavelength.

Aperture field measurements for circular apertures of intermediate size, $D=3\lambda$ to $D=8\lambda$, (that is, $ka=3\pi$ to $ka=8\pi$), carried out at a wavelength of 3.2 cm are shown in Fig. 6 where again the agreement with calculations derived from Eq. (1) is seen to be good. For the sake of comparison with the exact theory of Meixner and Andrejewski,[17] the latter author's calculations for an aperture $ka=10$ were used as an additional check on Eq. (1), although no experimental results were made on an aperture of exactly this size; however, measurements on the aperture $ka=3\pi$ shown in Fig. 6 could be used. The comparison between the two theories is depicted in Fig. 7. A plot of the amplitude of the electric field at the center of a circular aperture as function of the aperture size (ka) appears in Fig. 8. Results from the exact theory mentioned above are shown by a heavy line, those from Eq. (1) by a broken line and experimental data by solid dots. As in the previous case, Eq. (1) fails when $ka<3$; that is, when the aperture diameter is less than about one wavelength.

Experiments on the very largest aperture $D=40\lambda$ were done at a wavelength of 1.25 cm. No attempt was made in this case, or in any of the work which follows, to normalize the intensity to that of the incident wave.[26]

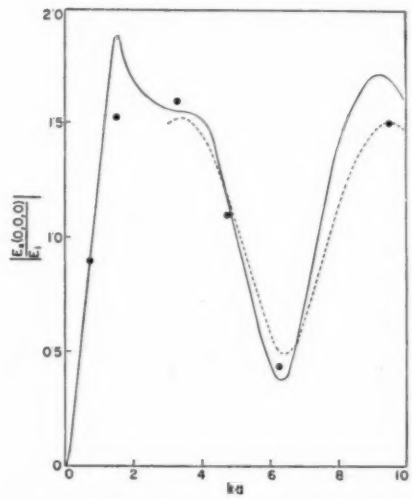


FIG. 8. The electric field at the center of a circular aperture as a function of the aperture size (ka). [●● measured; —— calculated from the exact theory of Meixner and Andrejewski; - - - - calculated from Eq. (1).]

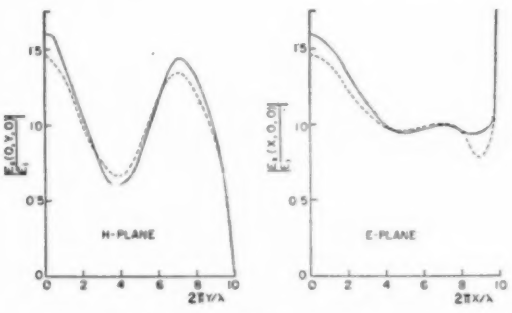


FIG. 7. The electric field within a circular aperture, $ka=10(D=3.18\lambda)$. E- and H-plane distributions. [—— calculated from the exact theory of Meixner and Andrejewski; - - - - calculated from Eq. (1).]

In order to investigate the generality of Eq. (1) under conditions other than plane wave illumination, a small source was placed at finite distances from the plane of the screen thus illuminating it by a non-homogeneous spherical wave. Results of H-plane measurements only are shown (Fig. 9) for various distances (z_0) of the source from the aperture, with the source located on the axis of symmetry z. Patterns for the three distances z_0, namely, $z_0=100\lambda$, 89λ, and 80λ are given under captions A, B, and C of Fig. 9. Here the experimental arrangement was such that the taper in the amplitude of the illuminating wave at the edge of the aperture as compared with that at the center was -3 decibels. The pattern marked D, measured with

[25] C. J. Bouwkamp, Philips Research Repts. 5, 401 (1950).

[26] The normalizing constant of the decibel scale employed henceforth is arbitrary and the level of calculated values is adjusted to result in the best possible fit with experimental data.

the source at $z_0=80\lambda$ differed from pattern C in that the source was arranged to give a taper of only -0.5 decibels from center to edge. The difference in the two patterns (C and D), which resulted from a difference in the amplitude distribution alone (the phase distribution remaining unaltered), should be noted. The reduction in the aforementioned taper (accompanied by a negligible change in the phase distribution) was achieved by replacing a $2\frac{1}{2}$ by 3 wavelengths rectangular horn, which was used as the source, with an open wave guide.

Calculations (open circles) again were made from Eq. (1)[27] taking into account the phase and amplitude distribution of the incident wave. The agreement with measured values is seen to be good. Similar patterns,

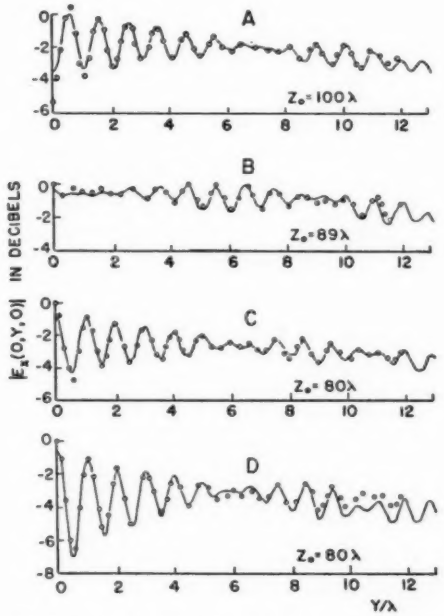


FIG. 9. Amplitude of the electric field within a circular aperture, $D=40\lambda$, illuminated by an inhomogeneous spherical wave. H-plane distribution. [—— measured; ○○ calculated from Eq. (1).]

showing the effect of the varying phase distribution upon the envelope of the actual diffraction pattern were observed earlier by Silver and Ehrlich[14] but were not explained quantitatively.

In addition to the aforementioned aperture field measurements near-zone patterns were also obtained for distances z chosen at random between zero and ten wavelengths. In Figs. 10 and 11 are shown such measurements for two apertures of intermediate size $D=7.5\lambda$ and $D=8\lambda$. Due to the difficulty of evaluating the integral in Eq. (1) for $z>0$, a paraxial approximation was made and was used for comparison.[8,15] The calculations are only good for displacements x, y from the axis of symmetry z not greater than about $D/4$.

[27] In D. C. Hogg's original work [J. Appl. Phys. 24, 110 (1953)] calculations were made from Andrew's semiempirical formula [J. Appl. Phys. 21, 761 (1950)] modified to allow for phase and amplitude variations in the irradiating wave. However, it can be shown that the formula used is derivable from our Eq. (1).

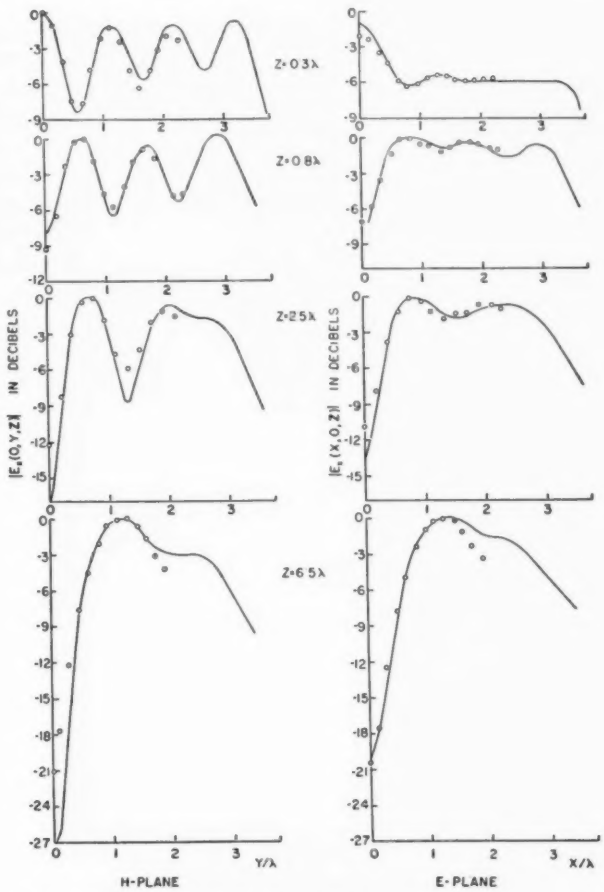


FIG. 10. The electric field in the near-zone of a circular aperture, $D=7.5\lambda$, at various distances z from the plane of the screen. E- and H-plane distributions. [—— measured; ○○ calculated from Eq. (1).]

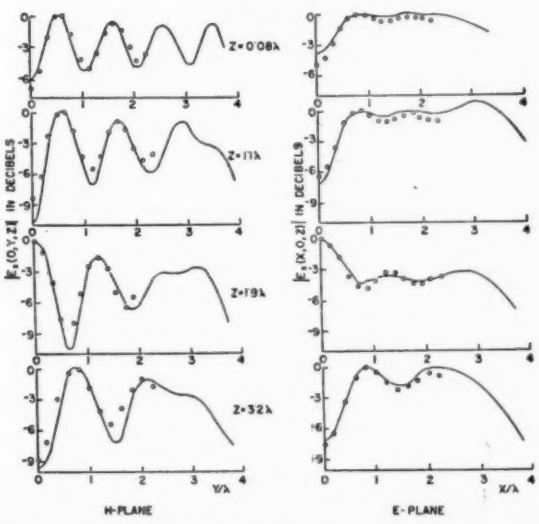


FIG. 11. The electric field in the near-zone of a circular aperture, $D=8.0\lambda$, at various distances z from the plane of the screen. [—— measured; ○○ calculated from Eq. (1)].

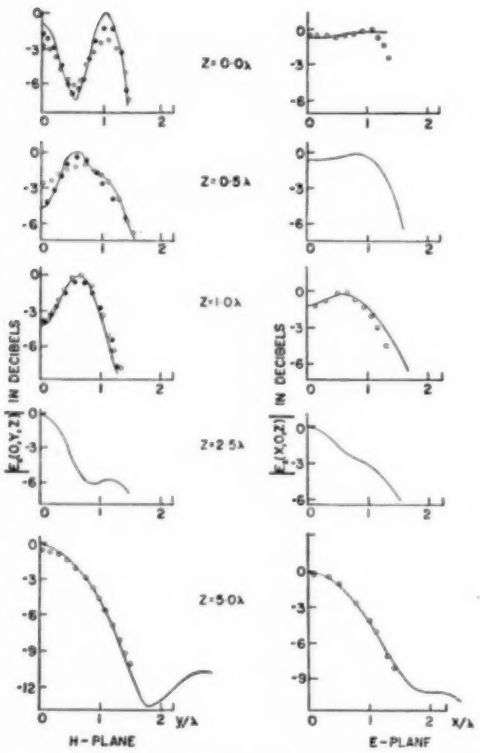


FIG. 12. Amplitude of the electric field within and in the near-zone of a long slit, $2a=3\lambda$, at various distances z from the plane of the screen. [—— measured; ○○ calculated from Eq. (1); ●● calculated from an empirical modification of Eq. (1).]

(b) The Infinite Slit

Measurements on long slits illuminated by plane waves were carried out at a wavelength of 3.2 cm. Preliminary tests showed that the slit which was about 40 wavelengths long, behaved as if it were infinite in that no change in any of the diffraction patterns was noted when the slit was foreshortened by as much as 10 wavelengths. The incident wave was plane over the slit's characteristic dimension (i.e., its width equal to $2a$) but was found to taper by about 2.5 decibels over its length; this taper did not affect the diffraction patterns to any measurable extent, as was found by subsidiary experiments.

The field was scanned in both the E and H planes. Here, a more careful definition of these planes is desirable; with the incident electric vector in the direction of the long side of the slit and oriented along the x-axis of the previously defined coordinate system, an H-plane measurement is a scan along the y-axis; an E-plane measurement is one in which the incident electric component, again pointing along the x-axis, is now oriented perpendicular to the length of the slit, the scan being made along the x-axis.

Figures 12–14 show measurements and calculations for three slits of widths $2a=3\lambda$, 6λ, and 7λ for a range of distances z varying from $z=0$ to $z=5\lambda$ in integral wavelength steps. Computations from Eq. (1) are shown by open circles. In general the agreement is satisfactory except that in certain cases marked discrepancies will be observed: in the H planes of Figs. 13 and 14 at a distance $z=\lambda$, the theoretically computed maxima and minima are seen to be out of step with the measured patterns; that is, the computed maxima fall on observed minima, and *vice versa*. A better agreement with experiment was found by an empirical reduction of the slit width by 0.25λ (solid

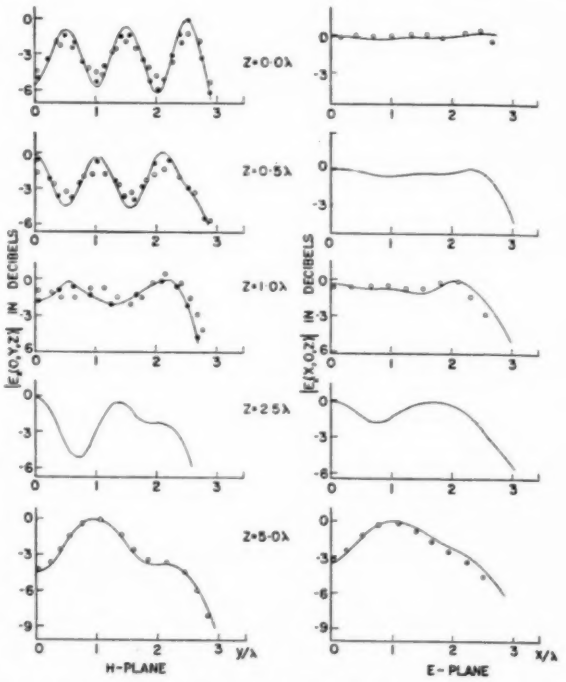


FIG. 13. The electric field within and in the near-zone of a long slit, $2a=6\lambda$, [—— measured; ○○ calculated from Eq. (1); ●● calculated from an empirical modification of Eq. (1).]

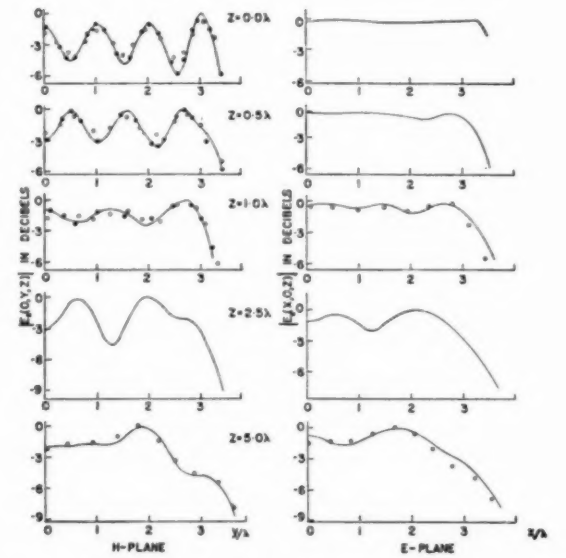


FIG. 14. The electric field within and in the near-zone of a long slit, $2a=7\lambda$. [—— measured; ○○ calculated from Eq. (1); ●● calculated from an empirical modification of Eq. (1).]

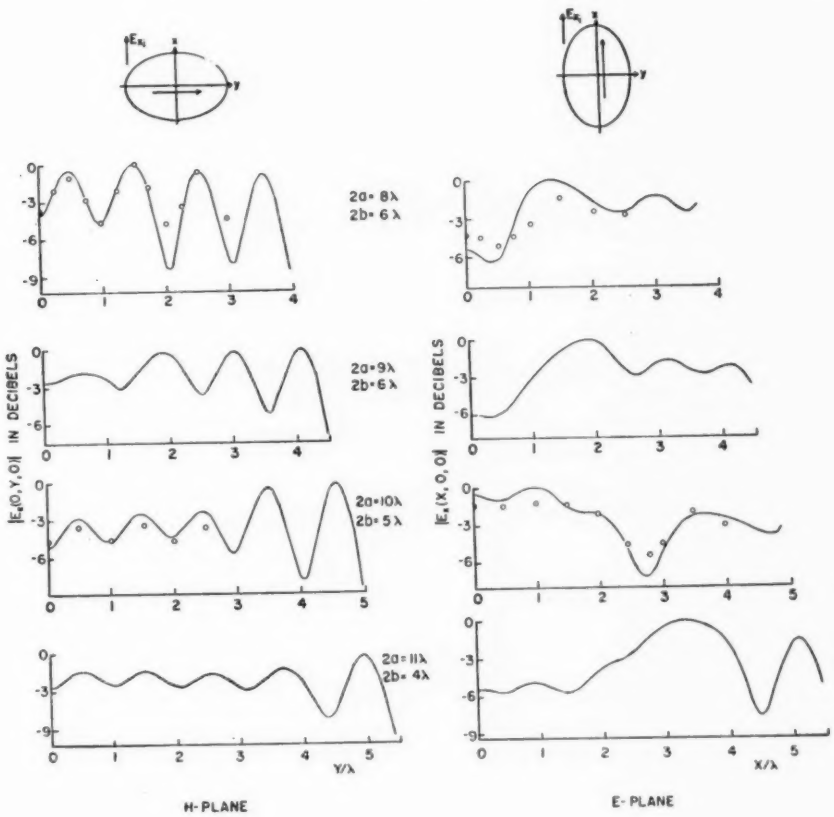


FIG. 15. Amplitude of the electric field within elliptical apertures of various sizes and shapes. Direction of scan (⟶) along the major diameter. [—— measured; ○○ calculated from Eq. (1).]

dots); the integration in Eq. (1) was not carried out over the actual slit width but over one which was one-quarter of a wavelength narrower. In patterns other than those mentioned above, a similar modification of Eq. (1) resulted in a somewhat better agreement with measurement than could be obtained from the theory itself. The most obvious discrepancies were generally met at those distances z behind the aperture which marked regions of transition from one type of field distribution to another. No theoretical justification can yet be given for this modification of Eq. (1).

(c) The Elliptical Aperture

The diffraction field of elliptical apertures of various eccentricities was examined in a manner analogous to that for the infinite slit. Here however, more stress was laid upon the aperture field, chiefly because of the difficulties encountered in evaluating the integral of Eq. (1) for field points outside the aperture plane. Four ellipses whose major ($2a$) and minor ($2b$) diameters had the following dimensions, were examined as to their patterns when illuminated by a plane wave:

(1) $2a=8\lambda$; $2b=6\lambda$

(2) $2a=9\lambda$; $2b=6\lambda$

(3) $2a=10\lambda$; $2b=5\lambda$

(4) $2a=11\lambda$; $2b=4\lambda$.

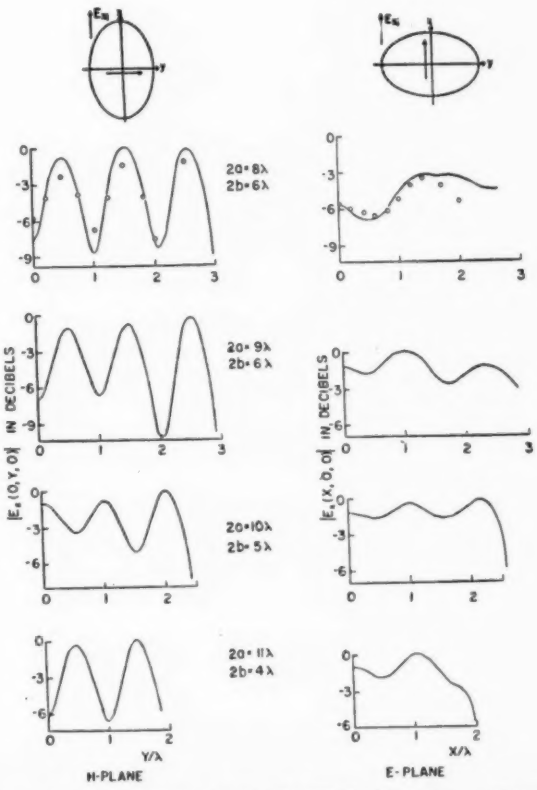


FIG. 16. The electric field within elliptical apertures. Direction of scan (⟶) along the minor diameter. [—— measured; ○○ calculated from Eq. (1).]

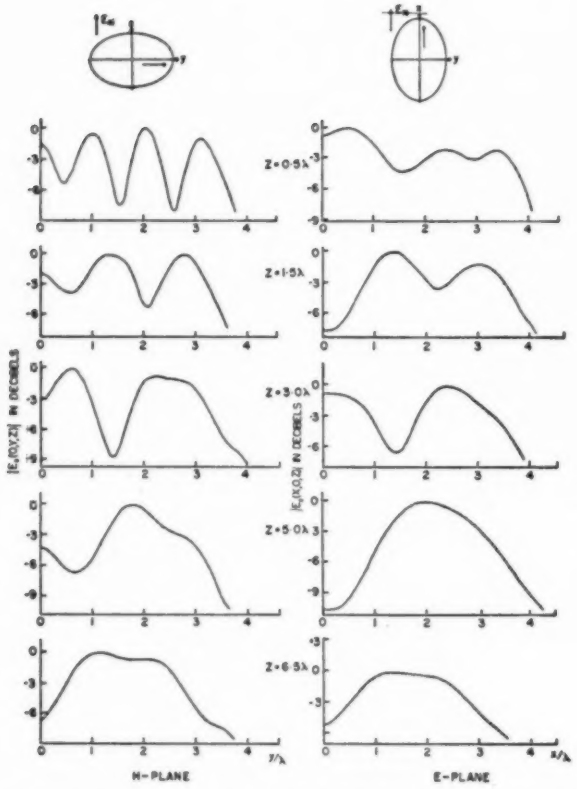


FIG. 17. The measured amplitude of the electric field in the near-zone of an elliptical aperture, $2a=8\lambda$, $2b=6\lambda$, at various distances z from the plane of the screen. Direction of scan (⟶) along the major diameter.

Two measurements in both the E and H planes were needed since the incident field could be oriented with its electric vector along either of the two principal axes of the ellipse. For the sake of clarity, diagrams illustrating the direction of polarization and direction of scan relative to the ellipse, are included in Figs. 15 and 16; these figures show some representative aperture-field measurements and their comparison with predictions from theory. Finally in Figs. 17 and 18 are depicted near field measurements ($z>0$) for a $2a=8\lambda$, $2b=6\lambda$ ellipse for which, however, no calculations are available.

(d) General Description of the Diffraction Patterns

So far the diffraction from each one of the apertures was discussed separately and no attempt was made to relate the observations on one aperture shape with those of the others. There are, however, several characteristics which can be integrated into a general picture.

A very striking property occurring in the H-plane aperture field patterns is the one-to-one correspondence between the characteristic aperture size and the number of maxima and minima in the intensity distribution. Suppose the characteristic dimension of the aperture [i.e., diameter ($2a$) in the case of a circle, slit width ($2a$) in the infinite slit and the major ($2a$) or minor ($2b$) diameters for the ellipse] are an integral number N of wavelengths in size. It is then found that there are N maxima and $(N-1)$ minima in the H-plane patterns. Also when N is even, a minimum appears at the center of the pattern ($x,y=0$) and a maximum when N is odd. This rule breaks down for very long elliptical apertures when measurements are made in the direction of the major diameter and in all those cases when the irradiating field is not plane in amplitude and phase over the face of the diffracting hole. If the aperture is a nonintegral number of wavelengths in size, the amplitude variations are generally found to be small and the rule no longer applies. The E-plane patterns usually exhibit much smaller and less regular variations in their amplitude than the corresponding H-plane measurements and no simple connection between the number of maxima and minima and the aperture size is evident.

An interesting point of contrast in the aperture field patterns for the different shapes of holes was also noted: while the H-plane measurements for the circular aperture exhibit their largest maxima (or minima) at

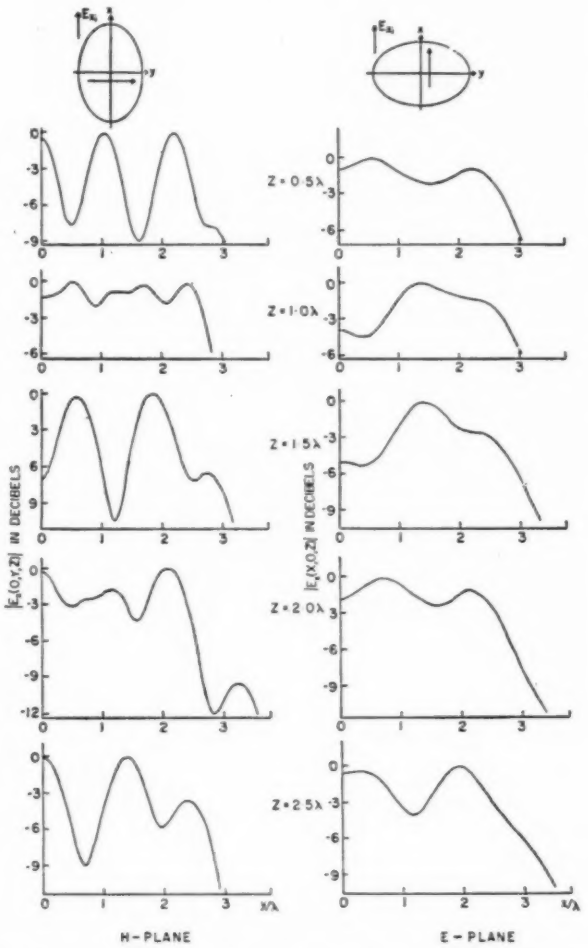


FIG. 18. The measured amplitude of the electric field in the near-zone of an elliptical aperture, $2a=8\lambda$, $2b=6\lambda$. Direction of scan (⟶) along the minor diameter.

the center of the hole, these appear in the case of ellipses in their focal regions and are absent for the infinite slit where the variations are sensibly uniform over the whole slit width.

In measurements made outside the aperture plane, one finds that with increasing distance z from the plane of the screen, the E- and H-plane patterns become progressively more similar to each other and at sufficiently large distances (greater than about 10λ) they become almost identical. It is in these regions that the classical Kirchoff-Fresnel theory (which is unable to take into account the polarization of the waves) predicts measurements with good accuracy and that is why the theory is found to be so successful in optical diffraction phenomena in which, of necessity, measurements must be made at considerable distances from the plane of the screen.

CONCLUSIONS

The theoretical and experimental work on the diffraction field of the circular aperture stimulated further research on the problem for apertures of other shapes. The results presented for the elliptical hole and the infinite slit are supplemented by an exhaustive investigation on circular apertures whose dimensions were varied over a range from one-quarter to forty wavelengths in diameter. In general, only the electric field was measured; in a few cases however, the magnetic component was also obtained.

Most of the measured field distributions were compared with predictions based upon a simple approximate theory. For the circular aperture of diameter equal to or greater than about one wavelength ($ka \geq \pi$) the agreement between theory and experiment was found satisfactory, irrespective of the nature of the illumination (plane or spherical wave). This agreement does not become substantially worse when the aperture shape is changed from circular to elliptical or to a long slit. Some discrepancies were noted in the near field of the slit in particular in the regions of transitions from one type of field distribution to another. An empirical modification of the theory (that is an effective contraction of the physical slit width) results in better agreement with experiment.

It is felt by the authors that sufficient parameters (a) the aperture shape, (b) the aperture size, (c) the nature of the illumination, were varied in the basic Eq. (1) to warrant the conclusion that, despite some noted discrepancies, the approximate theory predicts adequately the observed phenomena provided that the characteristic aperture dimension is not less than about one wavelength.

ACKNOWLEDGMENTS

The writers wish to acknowledge the valuable work of Miss E. Major who carried out most of the computations presented in this paper; they are also indebted to V. Avarlaid and his staff for their technical help.

Thanks are due the National Research Council of Canada through whose assistance to three of us (S. J. B., A. R. M., and D. C. H.) this work was made possible.

Announcements

Division of Fluid Dynamics

The Division of Fluid Dynamics of the American Physical Society has elected the following officers: S. Chandrasekhar, Yerkes Observatory, chairman; G. B. Schubauer, National Bureau of Standards, vice-chairman; and F. N. Frenkiel, Applied Physics Laboratory, The Johns Hopkins University, secretary-treasurer (he is the retiring chairman). The Executive Committee of the Division includes: J. C. Evvard, Lewis Flight Propulsion Laboratory, NACA; B. Lewis, Combustion and Explosive Research Inc.; H. W. Liepmann, California Institute of Technology; and S. A. Schaaf, University of California at Berkeley.

Noise Measurements in Semiconductors at Very Low Frequencies*

T. E. FIRLE AND H. WINSTON
Hughes Aircraft Company, Culver City, California
(Received January 10, 1955)

The low-frequency noise spectra of germanium and silicon rectifiers have been determined. Two methods were employed: a numerical analysis according to the Wiener-Khintchine relationship, and a frequency translation photo playback method. No evidence of a deviation from a $1/f^a (a>1)$ noise power law was found down to frequencies as low as 6×10^{-5} cps.

THE dependence of excess or flicker noise in semiconductors on frequency can be described over a large range of frequency by the relation

$$P_{exc} \propto f^{-a}, \quad (1)$$

where P_{exc} is the noise power per cycle, f is the frequency, and a is unity or a little greater. It is evident that this law cannot hold for frequencies approaching zero,[1] for then the integrated noise power would increase without limit. At some low frequency Eq. (1) is expected to break down, and the value of the frequency at which the relation begins to fail would be a guide to understanding the mechanism of excess noise. D. K. Baker,[2] investigating germanium rectifiers, found no break or "turnover" point down to frequencies as low as 0.002 cps. In the work reported here measurements were made at even lower frequencies on both germanium and silicon rectifiers.

It was found, as pointed out by Baker,[2] that extremely careful temperature control is necessary in order to avoid drifts due to changes in reverse current. The temperature control was achieved by immersing the rectifier under investigation in an oil bath. The oil was continuously agitated and the oil bath itself was mounted in an aquarium-type water bath of large heat capacity. The water circulated continuously and changes in temperature were sensed by a DeKhotinsky two column mercury thermo-regulator which controlled the heating elements. Temperature variations greater than 0.005°C resulted in correlations between the noise and temperature fluctuations, observed visually in some cases and by numerical analysis in others. We used a Beckman differential thermometer with telescope to measure temperature changes down to 0.001°C. When the temperature changed more than this amount in the course of a run, there was always a noticeable drift in the galvanometer deflections. Therefore, we used freedom from drift in the experimental records themselves as a criterion for the absence of temperature changes greater than 0.001°C. More than half of the two score runs made satisfied this criterion.

The rectifier under investigation was mounted as one arm of a Wheatstone bridge and the fluctuations about the balance condition recorded. Two different recording methods were employed in the course of the experiments: recording of the numerical values of the visually observed galvanometer deflections at equally spaced time intervals $\Delta\tau$, and photographic recording of the deflections on film.

The numerical data obtained by visual observation constituted the basis for a quantitative determination of the noise power spectrum according to the Wiener-Khintchine[3] relationship, which states that the autocorrelation function and power spectrum of a fluctuating variable are each other's Fourier cosine transforms. Standard IBM machine programs are available for determining the values of the autocorrelation function corresponding to correlation times ranging by steps of $\Delta\tau$ from zero up to 50 or 100 times $\Delta\tau$; there are also standard IBM methods for the Fourier inversion of the autocorrelation function given in such a form. The Computer Group at Hughes Aircraft Company under the direction of Essor Maso carried out the necessary IBM operations. Figure 1 shows the power spectrum determined from a single noise run on a germanium point contact rectifier; a 50-point program was used in the analysis.

Even with machine methods, computation of a 50 or 100 point program is a formidable task, to say nothing of the effort required to record the original numerical data. It became evident that more convenient recording and analyzing methods had to be

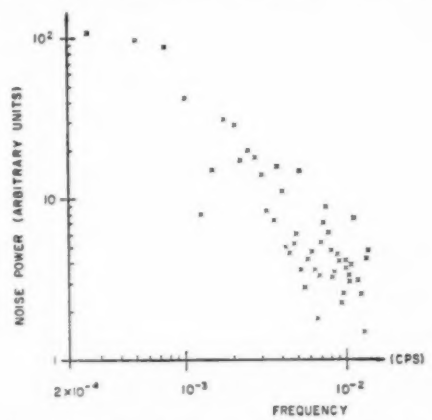


FIG. 1. Germanium point-contact diode, typical low-frequency noise power spectrum (machine computation).

* Supported in part by the Signal Corps.

[1] A. van der Ziel, *Noise* (Prentice-Hall, Inc., New York, 1954), p. 318.

[2] D. K. Baker, J. Appl. Phys. 25, 922 (1954). Ph. D. dissertation, University of Pennsylvania, 1953.

[3] See reference 1, p. 316.

found. The method finally adopted utilized a frequency translation of the low-frequency recording into the audio range so that standard audio analysis methods could be employed. Rollin and Templeton[4] have used this principle with magnetic tape recording. In our work a photographic film recording has been used instead.[5] The galvanometer position is recorded on a slowly moving film by one edge of a light beam reflected from the galvanometer mirror; the result is a variable area film recording of the galvanometer deflection. After photographic processing, the film is looped and played back at a greatly increased rate past a light beam shining on a photocathode, and the output of the phototube is applied to a conventional harmonic wave analyzer. Since the recording and playback speeds are known, the analysis of the low-frequency noise record can in this way be done in the audio range.

An analysis of an individual noise run of a silicon fused junction diode by means of the wave analyzer frequency translation method is shown in Fig. 2. The wave analyzer used gives voltage rather than power readings; therefore the slope of a log-log plot of voltage reading *vs* frequency must be multiplied by -2 to give a in Eq. (1).

Because of the statistical nature of noise, the results from any single run are expected to be quite erratic and could at best show only a general trend. If, however, the results of a large number of single runs under essentially identical conditions are superimposed on a log-log plot of noise voltage *vs* frequency, as shown in Fig. 3 for our data on a silicon junction rectifier, it is seen that the areas of greatest density of points lie along a straight line. It is quite clear that this line has a slope of magnitude greater than $\frac{1}{2}$ [corresponding to $a>1$ in Eq. (1)].

The lowest frequency which can be analyzed depends on the total observation time of the noise fluctuations, the lowest frequency at which the wave analyzer is designed to operate, and the available ratio of recording to playback speed. As an example, suppose the wave analyzer's lower limit is 30 cps and the total observation time 6 hours. Then, for an analysis down to 0.00006 cps, we need a speed-up ratio of 30 cps/0.00006 cps $=500\,000$.

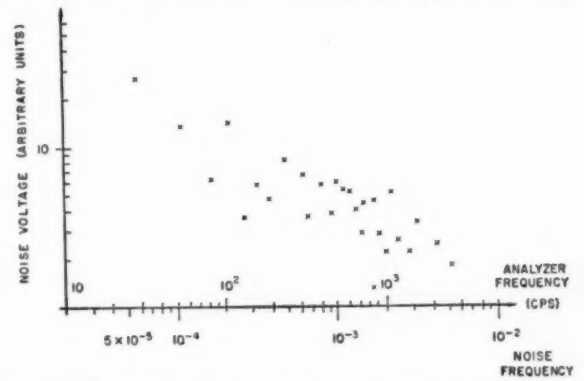


FIG. 2. Silicon *PN* junction diode, typical low-frequency noise analysis (photo playback method).

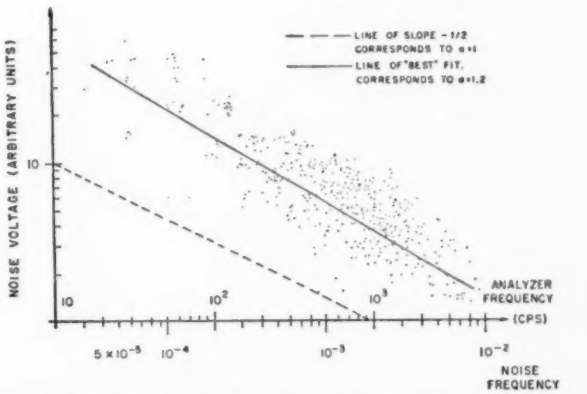


FIG. 3. Silicon *PN* junction diode, composite of 11 low-frequency noise analyses (photo playback method). The line of "best" fit was drawn in by eye; we do not wish to attach significance to the numerical value of a, except that it is greater than unity.

The germanium point contact diodes in our experiments were reverse biased to 4 volts; the reverse currents ranged from 5 to 20 microamperes. A back bias of 20 volts applied to the silicon *p-n* junctions gave reverse currents of 5 to 10 microamperes. The silicon diodes were selected for relatively high reverse currents and had a rather soft breakdown, most likely due to poor surface conditions. These diodes were actually chosen to exhibit high noise. The magnitude of the noise power at the low frequencies was calculated for some of the numerically analyzed runs. Both germanium and silicon rectifiers had noise levels at 0.001 cps about 120 db above thermal noise. Independent noise power measurements at 1000 cps gave noise levels above thermal noise of 40–50 db for both types of rectifiers used in this study.

One of the numerical runs was transcribed on film, so that the photo playback analysis could be compared with the machine analysis. The log-log plots of noise *vs* frequency for both methods could be superimposed on each other, indicating essential agreement between the two kinds of analysis. This comparison could have been used for calibrating the photo playback system to give the absolute magnitude of noise power, but we did not do this, since our major objective was to determine the frequency dependence.

Our results indicate no evidence of a deviation from the $1/f^a (a>1)$ excess noise power law for germanium point-contact diodes or silicon fused junction diodes down to 0.0005 cps and 0.00006 cps in frequency, respectively. The deviation of the excess noise power spectrum from its $1/f^a$ dependence must come at still lower frequencies. Experimental investigation of still lower frequencies than the ones reported here meets with great difficulties in keeping the parameters (particularly temperature) constant over the many hours required for the experimental runs.

[4] B. V. Rollin and I. M. Templeton, Proc. Phys. Soc. (London) **B66**, 259 (1953).

[5] A description of the apparatus is to be submitted for publication in another journal.

Calculation of the Space Charge, Electric Field, and Free Carrier Concentration at the Surface of a Semiconductor*

ROBERT H. KINGSTON† AND SIEGFRIED F. NEUSTADTER†
Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received November 2, 1954)

In the study of the surface properties of semiconductors, it is valuable to know the relation between carrier density at the surface and the shift in electrostatic potential from the bulk to the surface. By a solution of Poisson's equation utilizing Boltzmann statistics, the space charge, electric field, and change in free carrier concentration have been calculated for a semiconductor surface. The results are expressed as a function of the deviations of the Fermi energy from its intrinsic value in the bulk and at the surface. The calculated curves may be used for any nondegenerate semiconductor at any temperature, provided that the donor and acceptor levels are completely ionized. Numerical values are given for germanium and silicon at room temperature.

IN connection with surface studies on semiconductors at this laboratory, the need arose for the numerical relationships between space charge, carrier concentration, and electric field at a semiconductor surface as a function of the change in electrostatic potential from the bulk to the surface. These relationships cannot be determined explicitly, and a numerical integration was performed and the results obtained in graphical form. The calculations are a straightforward application of Shockley's treatment[1] and the numerical results are felt to be of value to those concerned with surface properties of semiconductors.

Figure 1 shows the energy band structure at the surface with ϕ considered positive if E_i lies below E_F, i.e., if the material is n-type. Expressions and graphs for three quantities will be determined:

(1) The space charge beneath the surface,
(2) The electric field at the surface, and
(3) The change in both electron and hole densities per unit area if ϕ at the surface changes from ϕ_B to ϕ_S.

In Poisson's equation

$$\partial^2\phi/\partial x^2 = -\rho/\kappa\epsilon_0;$$

the space charge, ρ, is given by

$$\rho = q(N_D - N_A + p - n),$$

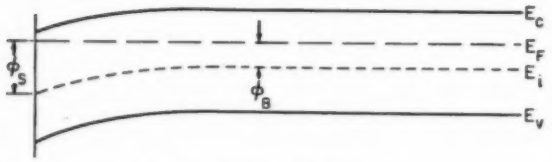


FIG. 1. Energy level diagram.

where N_D and N_A are the fixed ion densities,[2] and p and n are the free carrier densities. Now, in the bulk of the material, far from the surface, charge neutrality must exist. Therefore, for $\phi = \phi_S$,

$$N_D - N_A = n - p.$$

Writing the free carrier densities as

$$n = n_i e^{q\phi_B/\kappa T}, \quad p = n_i e^{-q\phi_B/\kappa T}$$

gives

$$N_D - N_A = 2n_i \sinh(q\phi_B/\kappa T),$$

and, in general, for any value of ϕ,

$$n - p = 2n_i \sinh(q\phi/\kappa T).$$

The resultant equation to be solved is therefore

$$\frac{\partial^2\phi}{\partial x^2} = -\frac{2qn_i}{\kappa\epsilon_0}\left(\sinh\frac{q\phi_B}{\kappa T} - \sinh\frac{q\phi}{\kappa T}\right).$$

If we convert to reduced notation, as follows,

$$u = \frac{q\phi}{\kappa T}, \quad L_D = \left(\frac{\kappa\epsilon_0 kT}{2q^2 n_i}\right)^{\frac{1}{2}},$$

the equation now reads

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{L_D^2}(\sinh u - \sinh u_B),$$

where L_D is the Debye length as defined by Shockley.[1] Integrating, from the bulk toward the surface,

$$\int_0^{\partial u/\partial x}\left(\frac{\partial u}{\partial x}\right)d\left(\frac{\partial u}{\partial x}\right) = \frac{1}{L_D^2}\int_{u_B}^{u}(\sinh u - \sinh u_B)du$$

* The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.
† Staff Member, Lincoln Laboratory, Massachusetts Institute of Technology.
[1] W. Shockley, Bell System Tech. J. **28**, 435 (1949).
[2] This assumes that the donors and acceptors are completely ionized, which is valid for silicon at room temperature and above and for germanium down to 100°K. For a general discussion, see W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand, New York, 1950), Chaps. 9 and 10.

gives the final result,

$$\frac{\partial u}{\partial x}=\pm\frac{\sqrt{2}}{L_D}[\sinh u_B(u_B-u)-(\cosh u_B-\cosh u)]^{\frac{1}{2}}. \quad (1)$$

To determine the value of the field at the surface, we let $u=u_S$;

$$E_s=-\left(\frac{\partial\phi}{\partial x}\right)_{\phi=\phi_S}=\left(\frac{\kappa T}{qL_D}\right)F(u_S,u_B),$$

where we put

$$F(u_s,u_B)=\sqrt{2}[\sinh u_B(u_B-u_s)-(\cosh u_B-\cosh u_s)]^{\frac{1}{2}}.$$

Similarly, by Gauss' law the space charge required to produce this field is

$$Q_s=\kappa\epsilon_0 E_s=\kappa\epsilon_0\left(\frac{\kappa T}{qL_D}\right)F(u_S,u_B).$$

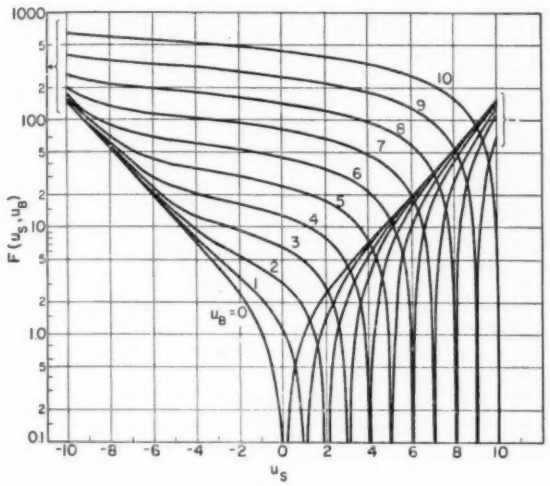


FIG. 2. $F(u_S,u_B)$ *vs* u_S for various values of u_B.

These expressions for the charge and the field may be rewritten, after substitution for L_D, as

$$Q_s=2qn_iL_DF(u_S,u_B)$$

and

$$E_s=\frac{2qn_iL_D}{\kappa\epsilon_0}F(u_S,u_B).$$

To determine the change in holes and electrons per unit area if the u at the surface is shifted from u_B to a final value u_S, it is necessary to evaluate the following two expressions:

$$\Delta P=n_i\int_0^\infty(e^{-u}-e^{-u_B})dx,$$

$$\Delta N=n_i\int_0^\infty(e^{u}-e^{u_B})dx.$$

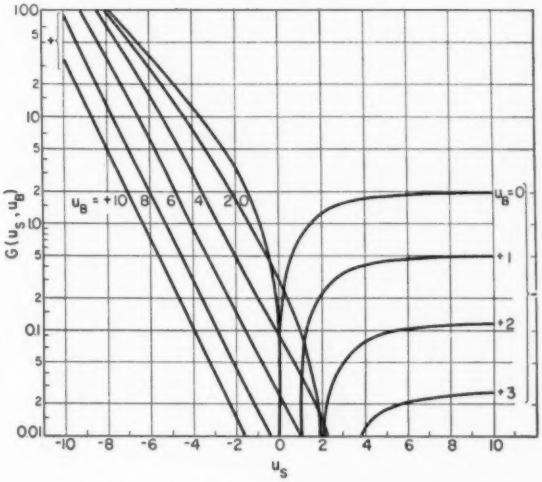


FIG. 3. $G(u_S,u_B)$ *vs* u_S for various values of u_B.

Considering the expression for the hole density and substituting from Eq. (1) gives

$$\Delta P=n_iL_D\int_{u_S}^{u_B}\frac{(e^{-u}-e^{-u_B})}{F(u,u_B)}du,$$

which is rewritten

$$\Delta P=n_iL_DG(u_S,u_B)$$

and thus describes the function, G. Symmetry considerations lead to the complementary form

$$\Delta N=n_iL_DG(-u_S,-u_B),$$

giving the change in electron density at the surface.

The functions, F and G, have been plotted in Figs. 2–4, for integral values of u_B with u_S as the independent parameter. Also shown on the charts is the sign of the function, which, if positive for F, gives positive space charge and electric field pointing from the bulk to the surface; if positive for G, gives an increase in

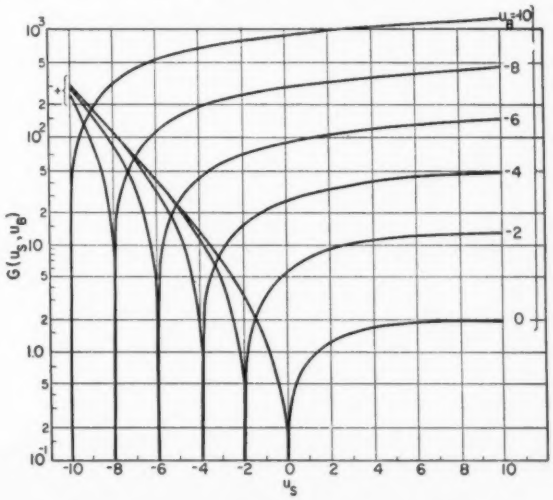


FIG. 4. $G(u_S,u_B)$ *vs* u_S for various values of u_B.

TABLE I. Numerical conversion values for germanium and silicon at 25°C.

	Germanium	Silicon
Q_S(coulombs/cm²)	$5.2\times10^{-10}F(u_S,u_B)$	$1.3\times10^{-11}F(u_S,u_B)$
E_S(volt/cm)	$370F(u_S,u_B)$	$12F(u_S,u_B)$
ΔP(holes/cm²)	$1.6\times10^{9}G(u_S,u_B)$	$4.2\times10^{7}G(u_S,u_B)$
ΔN(electrons/cm²)	$1.6\times10^{9}G(-u_S,-u_B)$	$4.2\times10^{7}G(-u_S,-u_B)$

the carrier density considered. The pertinent formulas are summarized below with numerical values of the coefficients for germanium at 25°C. For values of u_B less than zero, the additional identity for F may be derived by symmetry considerations.

$$Q_s=2qn_iL_DF(u_S,u_B),$$

$$F(u_S,u_B)=-F(-u_S,-u_B),$$

$$E_s=\frac{2qniL_D}{\kappa\epsilon_0}F(u_S,u_B),$$

$$\Delta P=niL_DG(u_S,u_B),$$

$$\Delta N=n_iL_DG(-u_S,-u_B).$$

In Table I the pertinent formulas are given for germanium and silicon at 25°C. We wish to thank Gertrude J. Durfee for the numerical calculations.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 6 JUNE, 1955

Electrical Breakdown in High Vacuum

W. S. BOYLE, P. KISLIUK, AND L. H. GERMER
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received September 9, 1954)

Currents preceding breakdown have been measured between closely spaced tungsten electrodes in high vacuum. It is found that field emission currents sufficient to evaporate anode metal flow before breakdown. These currents follow the Fowler-Nordheim equation when field magnification due to surface irregularities on the cathode is taken into account. The field magnification is a function of distance at electrode separations less than 4×10^{-4} cm.

Explanation of the observed breakdown at low voltage and small spacing requires an unusually high yield of electrons at the cathode per ion formed in the gap. Furthermore there is no measurable direct enhancement of the current by ionization even at higher voltages. The high electron yield must therefore exist over the entire observed range of breakdown voltages. This high yield is satisfactorily accounted for by the increase in field emission due to the positive ion space charge, which in turn increases the positive ion current density until there is breakdown. It is shown that breakdown occurs when the field emission current is increased by only 65 percent. This condition is reached with the ion current density much smaller than the electron current density.

INTRODUCTION

THE electrical breakdown between plane parallel clean metal surfaces in high vacuum has been investigated for gaps shorter then 8×10^{-4} cm, where the breakdown voltage is less then 2.2 kilovolts. While such short gaps may be of little practical importance in themselves, these experiments clarify the physical principles involved and are important in understanding breakdowns that occur between electrical contacts at low voltages in air.

It is shown that, for the majority of the breakdowns, field emission currents from small irregularities on the cathode evaporate metal from the anode, and that breakdown occurs in the resulting vapor. Occasionally, for very small irregularities, the cathode may be evaporated by Joule heating because of the high current density in the metal, as has been observed by Dyke and Trolan[1] for the very special geometry of the field emission microscope. For short gaps, however, the vapor is less likely to arise from the cathode than from the anode, since it can be shown for most of the irregularities which occur that the power density arriving at the anode surface is much greater than that dissipated in the region beneath the cathode surface. Furthermore, the "tilt," observed by Dyke and Trolan caused by increasing emission as the cathode is heated, is never observed in our experiments.

Short gap breakdowns in metal vapor, unlike the usual Townsend breakdown, occur even when the multiplicaion of the electron current by ionization in the gas is quite small. This is made possible by the high yield of electrons at the cathode per positive ion created in the gap. This high "γ" is caused by the high field present at the cathode at breakdown. It is not, however, a positive ion "γ" process in the usual sense, in that the collective space charge of the ions is responsible for the increased emission, rather than the processes associated with the impact of a single ion on the cathode.

[1] W. P. Dyke and J. K. Trolan, Phys. Rev. **89**, 799 (1953).

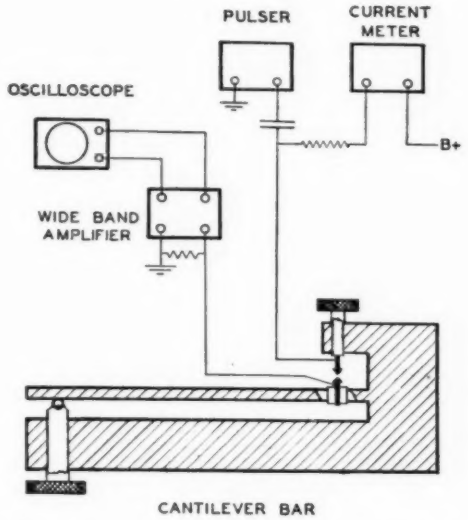


FIG. 1. Block diagram of equipment for measurement of currents between closely spaced electrodes.

EXPERIMENTAL PROCEDURE

In all of the experiments reported here the electrodes consist of a pair of crossed tungsten wires 0.75 mm in diameter. One electrode is fixed to a glass cantilever bar which gives a mechanical reduction of 60. A micrometer acting through a sylphon bellows at the end of the cantilever bar allows the electrode separation to be adjusted to within 200 A. A thorough baking of the vacuum system at 400°C and a further outgassing of the electrodes at 1800°C gave an operating pressure of 10^{-9} mm Hg. By using a blocking condenser, either steady or pulse voltages or both could be applied to the electrodes. Pulsed prebreakdown currents were displayed on an oscilloscope through a wide band amplifier with a rise time of 7×10^{-9} second. A block diagram of the system is shown in Fig. 1.

FIELD EMISSION CURRENTS

Steady field emission currents up to about 100 microamperes and pulsed currents up to several milliamperes could be drawn across the gap. Oscilloscope traces of the voltage pulse and the resulting field emission current pulse are shown in Fig. 2. That these are indeed field emission currents follows directly from the strong voltage dependence and the agreement with the Fowler-Nordheim equation[2] to be shown below.

FIG. 2. Oscilloscope trace of 1 μsec voltage pulse (left) and resulting field emission current pulse (right).

[2] R. H. Fowler and L. W. Nordheim, Proc. Roy. Soc. (London) A119, 173 (1928).

It was not possible under controlled conditions to carry the direct currents up to as large a value as the pulsed currents because of the bulk heating of the electrodes and the resulting change in the electrode separation.

Figure 3 shows field emission currents at an electrode separation of 19 000 A plotted against $1/V$. The curve follows smoothly from the dc to the pulsed region with some evidence of space charge saturation at the highest currents. From the linearity of the curve one must conclude that most of the current is being drawn from a single surface irregularity. If two or more separate regions were effective the curve would be concave upwards with the region with large emitting area effective at low voltages and the region with large field multiplication effective at large voltages.

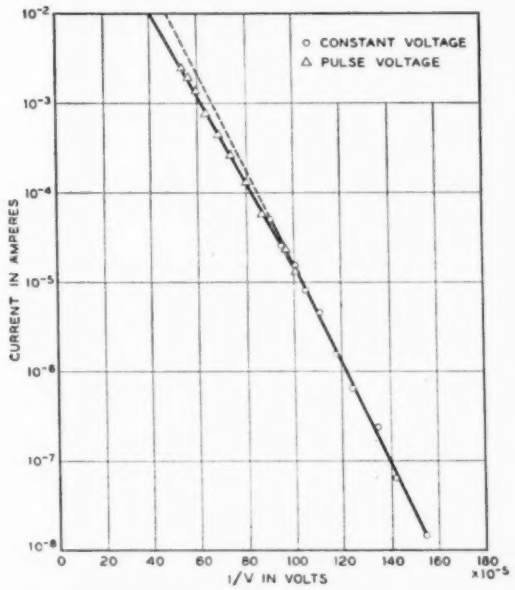


FIG. 3. Logarithm of current *vs* $1/V$ for contacts spaced at 19000 A. This curve yields $\beta=11$, $A=3\times10^{-11}$ cm² for $\Phi=5$ volts.

From the slope of the curve, using an approximate value of 5 volts for the work function of the tungsten surface, one finds that the field multiplication by irregularities on the surface is about 11 for this electrode separation. This may be obtained directly from the Fowler-Nordheim equation. This yields the true field at the surface, and from it the effective emitting area can be calculated. In Table I are presented, for each of four different surface irregularities, the field multiplication, β, at the surface and the emitting area as these are calculated.

It is evident that the field multiplication for any particular surface irregularity must decrease with decreasing electrode separation, when the separation is measured from the anode surface to the tip of the irregularity. At some small electrode separation where the radius of curvature of the surface irregularity is

TABLE I. Examples of field multiplication and emitting areas of surface irregularities.

	Distance	Field mult. factor, β	Emitting area, cm²	Diameter of circular area
No. 1	9000 A	13	1.7×10^{-13}	50 A
No. 2	24 500 A	14	1.6×10^{-10}	1400 A
No. 3	44 500 A	22	7.7×10^{-13}	100 A
No. 4	19 000 A	11	3.0×10^{-11}	600 A

much larger than the electrode separation the field multiplication must be unity.

In order to follow the variation of β with distance, the potential to draw some fixed small field emission current (10^{-8} ampere) was measured as a function of electrode separation. From this one can calculate as a function of distance the apparent field necessary to draw this current from a particular regularity. Using an effective emitting area obtained from a Fowler-Nordheim plot taken at one of these distances, the true field required to draw this current is determined. The field multiplication, β, follows directly. Curves for these experimental observations are presented in Figs. 4 and 5.

MAXIMUM PREBREAKDOWN CURRENTS

It was found that at any particular electrode spacing the maximum pulse current before breakdown varied between certain limits. These limits differed by a factor of about five. In order to smooth out this random variation, the average of 100 maximum prebreakdown currents was determined for each particular value of the pulse voltage over a range of pulse voltages. Experimentally it was more convenient to set the pulse voltage at a known value and then close up the electrodes, rather than to set the electrode separation and then raise the pulse voltage. These averages of the maximum prebreakdown currents are presented in Fig. 6 as a function of pulse voltage. Making use of the known approximate variation of the field intensification at the surface with distance, the maximum prebreakdown current as a function of distance can be calculated.

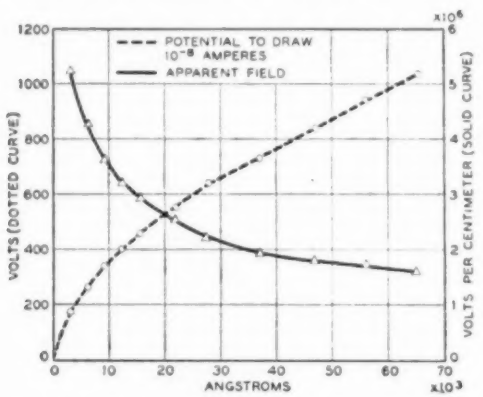


FIG. 4. Voltage and apparent field required to draw a field emission current of 10^{-8} amp plotted against electrode spacing from a particular cathode irregularity.

The breakdown voltage and the breakdown current are presented as a function of distance in Fig. 7.

The increase in apparent field at the shortest distances in Fig. 7 is a consequence of the decrease in the field multiplication of surface irregularities at small distances. The observed increase in maximum prebreakdown currents with distance has little effect on the breakdown field because of the extreme sensitivity of the field emission currents to small changes in field. Thus breakdown fields increase at very small distances for practically all metal surfaces, since no known means of polishing metal removes the submicroscopic irregularities.[3]

THE PRESENCE OF METAL VAPOR IN THE GAP

Experimentally the maximum prebreakdown current increases linearly with distance, and the breakdown voltage increases in a rather similar fashion. It follows, if one assumes linear spreading of the electron beam from the emitting area on the cathode, that some critical power flux at the anode determines the point of initiation of the breakdown. Calculation of the actual power flux requires a knowledge of the electron trajectories, which, in turn, necessitates the choice of a model for the cathode surface irregularity giving rise to the emission. It will be shown that after the heat conducted away from the anode spot has been taken into account, the remaining power is sufficient to create a considerable vapor density. The breakdown is then, in fact, a gas breakdown.

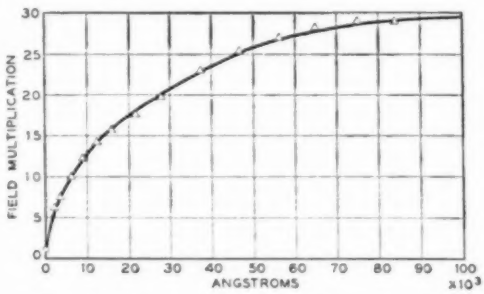


FIG. 5. Field multiplication *vs* electrode spacing. The points are derived from the constant current plot and are for a particular cathode irregularity.

Consider, for example, a breakdown at two kilovolts. The gap is then 7.5×10^{-4} cm and the average maximum prebreakdown power is 26 watts. Let us compute the size of a flat circular disk on the anode sufficiently large to conduct this power away in a steady state when the surface is at the boiling point. The power carried away by thermal conduction is $P_c=4KT_B\rho$ where K is the thermal conductivity, T_B the boiling temperature, and ρ the radius of the anode disk. Using the appropriate values for tungsten, one finds that to carry away 26

[3] C. G. Morgan and D. Harcombe, Proc. Phys. Soc. (London) **B66**, 665 (1953); F. L. Jones and C. G. Morgan, Proc. Roy. Soc. (London) **A218**, 88 (1953).

watts the radius of the anode disk would have to be $\rho=1.1\times10^{-3}$ cm. The electron beam would have to diverge at an angle of 56° to the normal to the cathode surface. This angle is much larger than can be attributed to any surface irregularity which yields the observed curve of field multiplication *vs* distance (Fig. 5). One is thus led to conclude that this situation is impossible and that an appreciable fraction of the anode power is used in evaporating anode material. If the spreading of the electron beam is 20°, which is still quite large for reasonable shapes of the cathode irregularity, one has a radius of 2.7×10^{-4} cm at the anode, and only 6.5 watts can be conducted away. The pressure of evaporated metal at the anode, p_1, can then be computed from the formula[4]

$$P_E=3.5\sigma\rho^2p_1(m/RT_B)^{\frac{1}{2}}, \qquad (1)$$

where P_E is the power used in evaporating metal, σ is the energy necessary to evaporate one gram, and m is the atomic mass. This yields a pressure p_1 of 670 atmospheres at the boiling point, or a density about 35 times that at NTP. Of course, under such circumstances the use of free space electron trajectories is not valid, and the beam would spread by more than 20°.

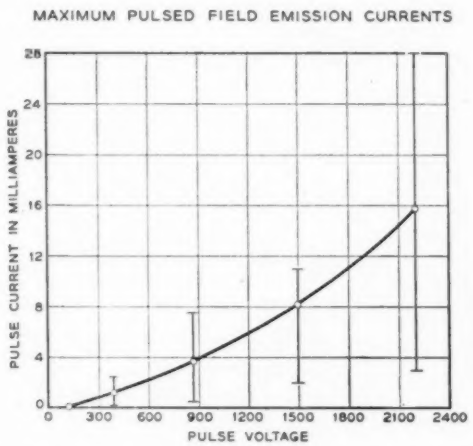


FIG. 6. Maximum prebreakdown current *vs* pulse voltage.

It appears that the prebreakdown vapor density is not easily susceptible to calculation, but must surely be considerable.

CATHODE PROCESSES DURING BREAKDOWN

Electrical breakdown under these conditions is the result of gas multiplication in the metal vapor and a secondary process of enhanced electron emission at the cathode due to the presence of positive ions (an effective γ_i). It is easy to show that the γ_i encountered in these breakdowns is unusually large (>1). We know in the first place that the number of ionizing collisions per electron is very small because of the low voltages at which breakdown occurs at very short distances, and this is shown also directly by an experiment where the pulse voltage is increased slowly right up to breakdown without measurable evidence from the resulting prebreakdown currents of gas multiplication (Fig. 8). The small ionization probability requires a large γ_i.

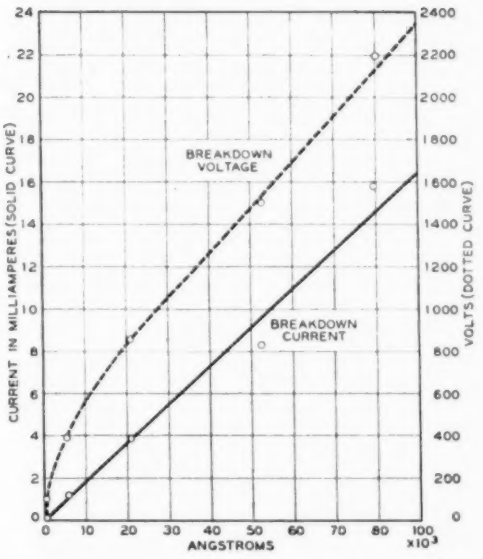


FIG. 7. Maximum prebreakdown current (——) and breakdown voltage (– – –) *vs* electrode spacing.

In the presence of a high field at the cathode, an appreciable electron yield per ion may result from a single positive ion approaching the surface.[5] On the other hand, this process is limited by the brief time the ion spends within 10 A of the cathode, and by the space charge of the emitted electrons. The resulting γ_i is too small to account for the lowest observed breakdown potentials even under optimum conditions of gas multiplication. The large yield observed is more likely the result of the increase of the field at the cathode by the combined space charge of all of the ions in the gap. This is not the γ_i generally considered in a Townsend type of discharge, and does not yield a constant number of electrons per ion since for one additional ion the increase in the number of electrons depends strongly on the number of ions already present as well as on the applied field. This enhancement of electron emission by positive ion space charge is considered in the following calculations and is shown to explain satisfactorily the experimental observations.

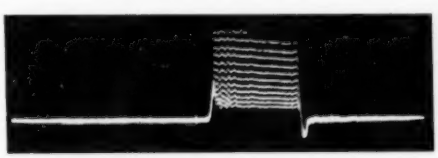


FIG. 8. Oscilloscope trace of 1 μsec prebreakdown pulse currents with uniform voltage increments.

[4] W. S. Boyle and L. H. Germer, J. Appl. Phys. **26**, 571 (1955), Eq. (2).

[5] R. R. Newton, Phys. Rev. **73**, 1122 (1948).

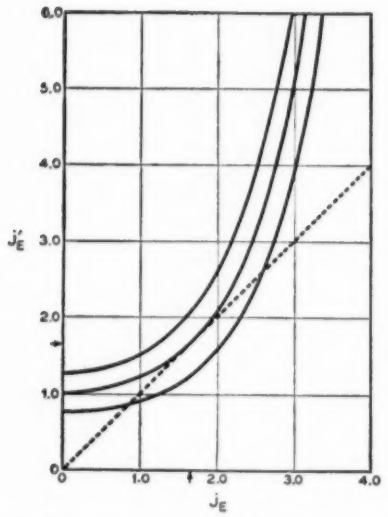


FIG. 9. Plot of $J_E'=j_E$(- - -) and $J_E'=j_0\exp(Mj_E^2)$ (——) for various values of j_0, in arbitrary units selected so that $j_0=1$ at breakdown.

The increment in field, E^+, caused by the positive ions is proportional to the ion current density, j^+, and we can therefore write

$$E^+=Cj^+. \tag{2}$$

To a first approximation, assuming the ions to be made near the anode and to proceed without collision to the cathode, the constant C can be written,

$$C=40\pi D(m/2e)^{\frac{1}{2}}/3V^{\frac{1}{2}}. \tag{3}$$

Here D is the gap length, V the voltage across the gap, m the mass of an ion, and e the electronic charge. Since, for a fixed gas pressure, the ion current density is proportional to the electron current density j_E from the cathode ($j^+=Gj_E$), we have $E^+=CGj_E$.

Using the simplified form of the field emission equation $j_E=A\exp[-B/E]$, and noting that the field at the cathode is the sum of the applied field, E_A, and that due to the ions, E^+, we have $j_E=A\exp[-B/(E_A+E^+)]$. The field due to the space charge of the electrons will be ignored.

Now, because of the extremely strong dependence of the electron current on the field, E^+ need be but a small fraction of E_A to increase the current greatly. $E^+\ll E_A$, and thus

$$j_E\simeq A\exp(-B/E_A)\exp(BCGj_E/E_A^2). \tag{4}$$

The gas multiplication factor, for such high values of E/P as exist in the gap before breakdown, is nearly proportional to the gas density. This in turn is proportional to the power density at the anode and thus to the electron current density at the cathode, if the emitting area is constant. Thus $G=\Gamma j_E$, where Γ is independent of j_E, as are also B and C. Furthermore E_A is only a very weak function of j_E. Replacing $BC\Gamma/E_A^2$ by a new constant M, Eq. (4) becomes

$$j_E=j_0\exp(Mj_E^2), \tag{5}$$

where $j_0=A\exp(-B/E_A)$ is the current density which would flow due to the applied field in the absence of space charge.

A graphical solution of Eq. (5) (Fig. 9) shows that, for small values of j_0, there are two solutions for j_E, of which only the smaller is stable. For large values of j_0 there is no solution and the breakdown condition has been exceeded. The breakdown condition is $\partial[j_0\times\exp(Mj_E^2)]/\partial j_E=1$, or

$$2Mj_Ej_0\exp(Mj_E^2)=1. \tag{6}$$

Equations (5) and (6) together give just at breakdown $j_E=(1/2M)^{\frac{1}{2}}=j_0e^{\frac{1}{2}}$. Thus at breakdown,

$$j_E/j_0=e^{\frac{1}{2}}=1.65. \tag{7}$$

Translated into words this states that breakdown takes place when the positive ion space charge in front of the cathode becomes sufficiently concentrated to increase the electron field emission by 65 percent over what the emission would be without this ion space charge.

An upper limit to the maximum positive ion current density which can flow before breakdown will be calculated, and from this an upper limit to the probability of an electron of the field emission current making an ionizing collision before it strikes the anode. A very rough calculation is sufficient. One first examines a field emission plot for a separation appropriate to breakdown at 800 volts. From this it is found that at a current near the upper limit, close to breakdown, an increase of the field emission current by the initial factor 1.65 corresponds to an increase of about 2×10^6 volts/cm in the field. This is the value of E^+ at breakdown. From Eq. (2) with a numerical value of the constant C from Eq. (3), one finds that this increase in field corresponds to a current density of positive ions j^+ of the order of 1.3×10^3 amp/cm^2. In this estimate the effect of field multiplication is neglected because we are interested only in an upper limit. Now the prebreakdown current j_E is of the order of 3×10^7 amp/cm^2 and therefore the ratio of ion to electron current densities $G=j^+/j_E$ is of the order of 4×10^{-5}. Even if we assume loss of ion current density by a factor of 10^3 due to spreading of the electron and ion beams, we still find as an extreme upper limit to the average number of ionizing collisions per electron the figure 0.04. It is clear that even at breakdown only a very small fraction of the electrons make ionizing collisions.

The conclusion from the above calculation is that the electrical breakdown of a gap between metal electrodes in high vacuum occurs when the current flowing between the electrodes exceeds by 65 percent the field emission current due to the *applied* field. This

increment of current is almost entirely additional field emission current elicited from the cathode by the enchanced field produced by the space charge of positive ions. Ionization of metal atoms vaporized from the anode produces these ions, but their number is so small that they make only an insignificant addition to the total current.

FORMATIVE TIME LAG

Evaporation of metal from the anode as calculated above is based on thermal conduction loss appropriate for a steady state. Until equilibrium is reached the conduction loss will be larger than for the steady state. It follows then that there will always be a thermal formative time lag which will depend on the amount by which the critical breakdown current (or power) is exceeded and on the thermal relaxation characteristic of the anode conduction process. This thermal formative time lag is not the formative time lag usually discussed, but is the time necessary to heat and evaporate enough anode material to give breakdown. The true formative time lag *which follows this* would be very difficult to calculate, since γ, the measure of the secondary process, is a function of the positive ion current density and will be increasing rapidly in the course of breakdown. Experimentally this true time lag is found to be less than 10^{-8} second and this is not surprising in light of the ion transit time, 10^{-11} second, at these small separations.

The time t for the surface of the anode to reach a temperature T for a given power input P is obtained from the relation,

$$T=(PT_B/P_c)[1-\exp(kt/a^2)\ \mathrm{erfc}(kt/a)^{\frac{1}{2}}] \qquad (8)$$

where P_c is the steady-state conduction loss at the boiling temperature T_B, k is the diffusivity, and s the radius of the area bombarded by the electron beam. The additional time lag required to build up the metal vapor pressure cannot be estimated since the breakdown pressure itself is not known. The average of a series of measurements of the formative time lag obtained from pulsed currents raised in small increments to breakdown (Fig. 8) yields a time of 0.3 μsec for a 10 percent current interval. The measurements were taken at 800 volts and at this potential a is about 5×10^{-5} cm so that, from Eq. (8), the temperature of the surface is at breakdown within 20 percent of the final equilibrium value which would be attained if there were no loss due to evaporation. The experimentally observed formative time lag and that calculated with these rather crude approximations are therefore in satisfactory agreement.

CONCLUSIONS

Field emission current can be drawn between closely spaced electrodes in high vacuum without breakdown. As the current is increased by increasing the field, anode metal is evaporated which gives rise to positive ions by electron collisions. Yet even up to breakdown there is no evidence of gas multiplication, and it can be shown that prior to breakdown the ion current density is only an extremely small fraction of the electron current density. After the field has been raised to a value high enough to cause breakdown, there is a formative time lag of the order of a microsecond without detectable current increase. This lag is the time necessary to heat and evaporate anode material. The actual final period of current increase is too short to be resolved on the apparatus used (less than 0.01 microsecond).

To account for breakdown it is necessary to conclude that the number of electrons released by each positive ion is quite large. It is shown that for the experimentally observed large breakdown fields the space charge of the positive ions does account for such a high electron yield, and furthermore that the current does not greatly increase before the breakdown condition is reached.

These vacuum arcs are closely related to the arcs discovered on closure of electrical contacts at low voltage by Germer *et al.*[6] The principal difference is that the breakdown at electrical contacts is facilitated by surface contamination on the electrodes and by molecules of air in the gap.

[6] L. H. Germer and F. E. Haworth, J. Appl. Phys. **20**, 1085 (1949); L. H. Germer, J. Appl. Phys. **22**, 955 (1951); L. H. Germer, J. Appl. Phys. **22**, 1133 (1951); L. H. Germer and J. L. Smith, J. Appl. Phys. **23**, 553 (1952); P. Kisliuk, J. Appl. Phys. **25**, 897 (1954); reference 4; M. M. Atalla, Bell System Tech. J. **32**, 1231 (1953); **32**, 1493 (1953); **33**, 535 (1954).

X-Ray Measurements of Particle Size and Strain Distribution in Cold Worked Silver

F. R. L. SCHOENING AND J. N. VAN NIEKERK
National Physical Laboratory, Council for Scientific and Industrial Research, Pretoria, South Africa
(Received November 29, 1954)

X-ray diffraction profiles from filed solid silver specimens were measured with a Geiger counter spectrometer using crystal-reflected Cu Kα radiation. It is shown that for values of $n>7$, the assumption of a Gauss strain distribution fits the observed Fourier coefficients far better than a Cauchy distribution. The results obtained from solid specimens are consistent with those previously measured for silver filings. The influence of background errors is discussed.

1. INTRODUCTION

THE interpretation of the Fourier coefficients, obtained from broadened x-ray diffraction lines, depends on the type of strain distribution which is assumed to be present in the material. This was pointed out in a recent communication by Williamson and Smallman,[1] who also showed that the results they obtained with iron and molybdenum filings could be interpreted more consistently by assuming a Cauchy rather than a Gauss strain distribution as first used by Warren and Averbach.[2]

The present authors[3] have recently investigated the broadening of the diffraction lines of silver filings and have interpreted the observations by assuming a Gauss strain distribution. These experiments have been repeated, using as specimens the filed surfaces of solid rods of high purity silver. Such solid specimens have the advantage that the geometrical surface conditions are better defined than for a powder sample, which is particularly important when the specimen surface forms part of the focusing arrangement as is the case with spectrometer recording. Furthermore, the same surface is used for recording the broadened and the sharp peaks.

These advantages seem to have eliminated certain difficulties previously encountered with the interpretation of the Fourier coefficients obtained from silver filings. On the other hand the lack of randomness and the appearance of definite recrystallization textures in the solid samples after annealing do not seem to limit the usefulness of the method, as good correlation was obtained between the results from both filings and solid silver specimens.

2. EXPERIMENTAL

Except for the following minor variations, the experimental procedure was the same as that described for the investigation on silver filings[3]: the x-ray observations were started immediately after filing the samples and the slit systems of the spectrometer were changed to fit the dimensions of the plane, filed surface (25×7 mm) of the solid specimens. The following experiments were carried out.

Sample *A*

A rod of high purity silver (99.999 percent) was filed under liquid oxygen to present a plane surface with the above dimensions. Immediately after filing, the 111, 200, 222, and 400 reflections were measured keeping the sample at −30°C. While exposing the sample to room temperature (23–26°C), the increase of the intensity maximum of the 200 reflection as a function of time was recorded. The sample was finally annealed at 300°C for one hour and the reflexions were again measured.

Sample *B*

A rod of high purity silver was filed at room temperature, dropped into liquid oxygen and transferred to the goniometer where the 111, 200, 222, and 400 reflections were measured while keeping the sample at −30°C. As for sample *A* the increase of the peak maximum of the 200 reflection and also the relevant profiles of the annealed specimen were again measured.

3. FOURIER ANALYSIS OF THE LINE SHAPES

Applying Stokes'[4] correction for instrumental broadening and using Lipson and Beevers strips, Fourier coefficients were calculated. Those obtained for sample *A* are shown in Fig. 1. The coefficients were interpreted:

(1) by assuming the strain distribution to be of the Gauss type (lnA_L plotted *versus* l_0^2),

(2) by assuming a Cauchy distribution (lnA_L plotted *versus* l_0).

The resulting values for A_L^P are shown in Figs. 2(a) and 2(b). From these diagrams it is evident that

TABLE I. Particle size and strain values obtained from solid specimens. Corresponding values from filings are given in brackets.

Sample	Particle size in A		rms strain %	
	111	100	111	100
A	230 (280)	180 (140)	0.40 (0.25)	0.40 (0.15)
B	270 (200)	150 (130)	0.37 (0.40)	0.28 (0.50)

[1] G. K. Williamson and R. E. Smallman, Acta Cryst. **7**, 574 (1954).
[2] B. E. Warren and B. L. Averbach, J. Appl. Phys. **21**, 595 (1950).
[3] F. R. L. Schoening and J. N. van Niekerk, Acta Metallurgica **3**, 10 (1955).
[4] A. R. Stokes, Proc. Phys. Soc. (London) **61**, 382 (1948).

the assumption of a Gauss strain distribution leads to simple curves, from which particle sizes as tabulated in Table I were obtained. On the other hand, the assumption of a Cauchy type of strain distribution leads to points on the graphs which are typical of those which would obtain if it is attempted to approximate a Gauss distribution by a Cauchy distribution. There seems to be one notable exception where similar types of curves were obtained [see Fig. 2(b)]. This, however, is probably because of the relatively small strain values present in this particular case.

From the foregoing considerations, it would appear that the Gauss function is a better approximation of the strain distribution in silver than the Cauchy function. Accordingly, the rms strain values were calculated by making use of the Gauss distribution only. These

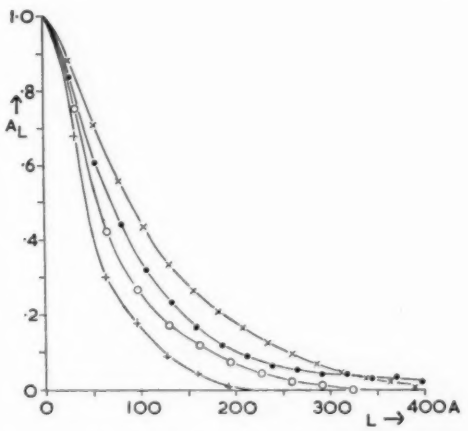


FIG. 1. Fourier coefficients calculated for solid silver specimens filed under liquid oxygen (sample *A*) x 111, ● 200, ○ 222, +400.

values, calculated from plots of $(\langle\epsilon^2\rangle)^{\frac{1}{2}}$ *versus* L, are tabulated in Table I.

4. RECRYSTALLIZATION AT ROOM TEMPERATURE

The increase of the 200 intensity maximum with time for the two samples when exposed to room temperature, is shown in Fig. 3. The existence of a definite incubation period, which is evident for one of the samples, suggests that recrystallization takes place in this specimen.

Further evidence that both samples recrystallize at room temperature is afforded by the observed changes in surface texture of the two samples. The intensity measurements listed in Table II reveal a definite change in surface texture of the samples after exposing them to room temperature. It is evident that for both samples the contribution from crystallites having their {111} planes parallel to the filed specimen surface decreases during annealing at room temperature while the contribution from crystallites with their {200} planes parallel to the surface increases considerably.

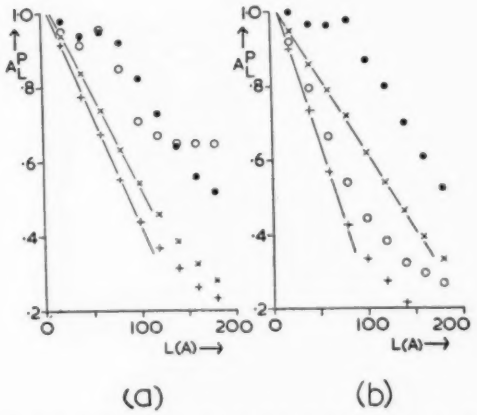


FIG. 2. (a) A_L^P for sample *A* (b) A_L^P for sample *B*. x 111, +100, assuming Gauss strain distribution. ● 111, ○ 100, assuming Cauchy strain distribution.

5. DISCUSSION

Comparing the results published in a previous communication on silver filings with those obtained from solid silver samples (see Table I), it is evident that agreement between corresponding particle size values is excellent. Strain values measured for the solid sample filed under liquid oxygen are higher than the values obtained from filings under corresponding conditions. The reason for this lies in the rapid recovery of such specimens when exposed to room temperature (see Fig. 3) as it must be remembered that, in the case of filings, the line profiles were only measured after the filings had been exposed to room temperature for one hour, the time necessary for sample preparation. In the case of solid silver specimens, cold worked at room temperature, it is evident from Fig. 3 that a relatively long incubation period is required before recrystallization starts. It is therefore to be expected that corresponding strain values measured for solid specimens immediately after filing and for filings after being exposed to room temperature for one hour should be comparable. One such pair of strain values (see Table I) confirms this. The discrepancy found in the other pair is probably due to the scattering of the points on the rms strain *versus* distance plots obtained for filings, which made the extrapolations uncertain. From this point of view the strain values obtained from the solid

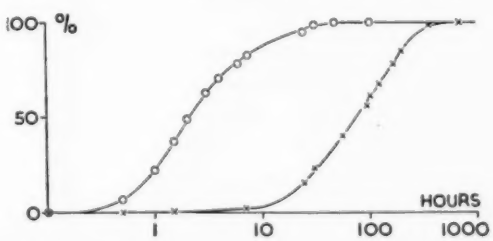


FIG. 3. Percentage increase of the counting rate of the 200 peak maximum at room temperature. Upper curve for sample filed under liquid oxygen, lower curve for sample filed at room temperature.

TABLE II. Integrated intensities of the 111 and 200 reflections. The intensities listed under each sample are expressed in terms of the 111 intensity measured immediately after filing and taken arbitrarily as unity.

	Sample A		Sample B	
	111	200	111	200
Immediately after filing	1.00	0.35	1.00	0.13
After 800 hours annealing at room temperature	0.73	1.32	0.74	0.51

specimens are much more reliable than corresponding values obtained from filings which must be considered as representing maximum values.

The lowest Fourier coefficients (Fig. 1) were calculated for L between 25 and 30 A, that is for n about 7. It is therefore only possible to conclude that the strain distribution in silver can be approximated better by a Gauss than by a Cauchy function for distances greater than 7 unit cells apart. The lower Fourier coefficients, in particular A_0, are severely influenced by background errors. Since A_0 is used for scaling all the higher coefficients, it was investigated what influence background errors amounting to 0.3 percent of the 111 peak maximum and 0.6 percent of the 222 peak maximum would have on the results. It was found that correction for such errors would make the straight line extrapolations of the $A_L{}^P$ *versus* L curves, obtained with the Gauss approximation, cut through $A_L{}^P=1$, but that it would have little influence on the shapes of the curves (derived from Gauss or Cauchy approximations) for $n>7$ and hence on the results obtained in the foregoing manner.

ACKNOWLEDGMENTS

The authors wish to thank Mr. F. T. Wybenga for assistance in accumulating experimental data and the South African Council for Scientific and Industrial Research for permission to publish this paper.

JOURNAL OF APPLIED PHYSICS VOLUME 26, NUMBER 6 JUNE, 1955

Energy Relations in Cold Working an Alloy at 78°K and at Room Temperature

J. S. L. LEACH,* E. G. LOEWEN, AND M. B. BEVER
Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received December 29, 1954)

The energy relations associated with the plastic deformation by cold working of a 75 percent gold–25 percent silver alloy at two temperatures were investigated. The samples were produced by drilling under controlled conditions, which is a suitable method of deformation. The energy stored in the chips was measured by tin solution calorimetry and the total energy expended in the deformation was determined from dynamometer measurements.

The results for cutting with a drill having the most suitable geometry were as follows. The energy stored at room temperature was 0.50 cal/g and the corresponding value for 78°K was 1.48 cal/g. The values of the total energy of deformation were 53.5 cal/g at room temperature and 109 cal/g at 78°K. The true stress-true strain curve at room temperature was normal for a nonferrous metal and the curve for 78°K showed the behavior expected at low temperature from a face-centered cubic metal.

INTRODUCTION

THE changes in the physical properties of a metal caused by cold working below room temperature are of considerable interest. Not only are electrical and mechanical properties different at different temperatures, but by cold working at low temperatures they are altered to a greater extent than by cold working at room temperature. Cold work causes changes in such properties as electrical resistivity by producing defects in the crystalline structure. Since these defects, such as dislocations, interstitial atoms and vacancies, also provide mechanisms by which a metal stores energy, it may be expected that a larger amount of energy is stored during cold working at low temperatures than at room temperature. The literature, however, contains little information on the energy stored in metals deformed at low temperatures.

In earlier phases of the research reported here, a calorimetric method for measuring the stored energy in a gold-silver alloy was developed.[1,2] It was shown also that machining by orthogonal cutting is a suitable method of deformation where measurements of the stored energy of cold work are to be made.[3] It has now been found possible to measure the energy stored in this alloy during deformation at 78°K by drilling under controlled conditions, which may be considered as related to orthogonal cutting. The energy stored in the chips at this temperature and at room temperature was measured calorimetrically. The total energy required to deform the metal by drilling at the two temperatures was also measured. Tensile tests were made on annealed samples of the alloy at the two temperatures.

* Present address: Murex Welding Processes Ltd., Waltham Cross, Herts., England.

[1] M. B. Bever and L. B. Ticknor, J. Appl. Phys. **22**, 1297 (1951).
[2] M. B. Bever and L. B. Ticknor, Acta Metallurgica **1**, 116 (1953).
[3] Bever, Marshall, and Ticknor, J. Appl. Phys. **24**, 1176 (1953).

EXPERIMENTAL PROCEDURES

Cold Working

Figure 1 shows the equipment used for cold working. The method consisted of drilling a block (A) of 75 percent gold–25 percent silver alloy which was rigidly mounted on a thermally insulating base (B). This base was in turn connected through a strain ring assembly[4] to the platform of a drill press. In this way the torque exerted by the drill could be measured and recorded. The mounting (C) of the block was designed to collect the chips in a small box (D) located under the alloy block.

Cuts were made at room temperature and at 78°K with straight-fluted and twist drills; the cutting face of the straight-fluted drill was ground radially so as to approach orthogonal cutting more closely. The drills were turned at 28 rpm and a feed of 0.005 in. per revolution was used. By cutting so slowly the local heating of the chip at the cutting edge was reduced. The room temperature cuts were made with the block and drill tip immersed in a small amount of carbon tetrachloride, which acted as a cutting fluid and cooling medium. During a cut the temperature of the carbon tetrachloride changed from about 25°C to about 31°C. The low-temperature cuts were made using a thermally insulating wall (E) to surround the alloy block and mounting, as shown in Fig. 1. The container formed by this wall and the thermally insulating base was filled with liquid nitrogen which maintained the block, drill tip, and chips at 78°K.

Four cuts could be made from each block of alloy. The drills (which ranged in size from $\frac{3}{8}$ in. to $\frac{23}{32}$ in.) and the thickness of the blocks were so chosen that the weight of each cut was about 12 grams. A pilot hole of $\frac{1}{8}$-in. diameter was machined in the block before cutting the first sample to avoid the extrusion effect normally occurring at the center of a drill.

For the calorimetric measurements it was necessary to know the weight of the charge. The chips produced at room temperature could be weighed directly, but weighing the chips cut at 78°K while preventing condensation or rise of temperature (neither of which could be tolerated), presented difficulties. These were circumvented by obtaining the weight of the chips added to the calorimeter as the difference between the weight of the alloy block before the cut was made and the weight of the block and any remaining chips after the calorimetric charge had been removed. Some cuts at room temperature, which were used as a check, confirmed the accuracy of this method.

The charge was kept at 78°K and free from condensation by leaving it immersed in liquid nitrogen until it could be prepared for adding to the calorimeter. The time of holding was about two hours.

[4] E. G. Loewen and M. C. Shaw, Instruments **23**, 6, 560 (1950).

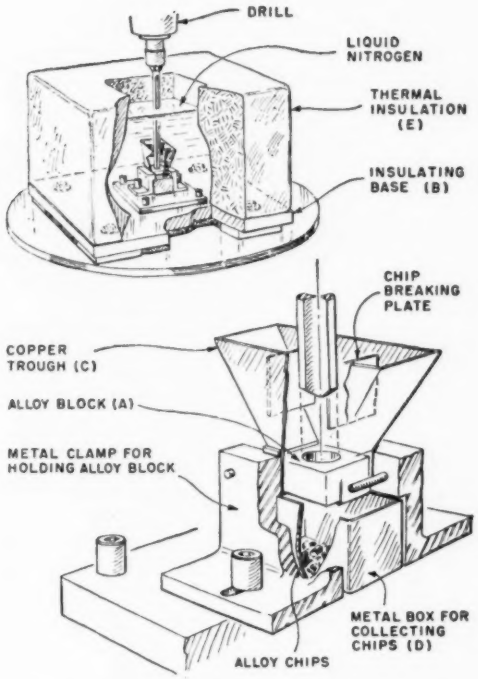


FIG. 1. Equipment for cold working by drilling under liquid nitrogen.

Calorimetry

The calorimetric equipment was that previously described[5] with some minor modifications. These modifications included improvements in vacuum and in temperature measurement and control. Also changes in the equipment made it possible to add a sample from the temperature of liquid nitrogen and to make additions without impairing the vacuum in the main part of the calorimeter.

In each run one cold worked and one annealed sample were added from 78°K to liquid tin at 240°C and the heat effects measured. The difference in these heat effects (adjusted for concentration[5]) represented the difference in the energy contents of the samples in their initial states, that is the stored energy of cold work at 78°K. Corresponding measurements were made starting at 0°C to determine the energy stored during drilling at room temperature.

The composition of the alloy was chosen in the earlier investigations[1–3] so that the energy required to heat a sample from 0°C to 240°C was just balanced by the heat evolution attending the dissolution of the sample in tin. Under these circumstances the change, due to the stored energy of cold work, in the measured heat effect could ideally be equal to and in practice was a large part of the total heat effect. This thermal compensation for any given composition is effective only over definite temperature intervals; thus the samples of the 75 percent gold alloy added from 78°K caused a relatively

[5] L. B. Ticknor and M. B. Bever, Trans. Am. Inst. Mining Met. Engrs. **194**, 941 (1952).

TABLE I. Total energy required in drilling experiments.

Type of drill	Room temperature in-lb/in.³	Room temperature cal/g	78°K in-lb/in.³	78°K cal/g
Twist drill	295 000	30.5	540 000	55.5
Straight-fluted drill	520 000	53.5	1 060 000	109

large heat effect. The accuracy to be expected from them is, therefore, less than from the samples of the same composition added from 0°C.

For the calorimetry it was necessary to obtain a clean dry sample without raising the temperature above 78°K. This was achieved by the following operations, during all of which the sample was kept immersed in liquid nitrogen. After drilling, the chips in the small box (D) (Fig. 1) were rinsed in fresh liquid nitrogen to remove any ice which may have condensed during the cutting process. The sample, now clean, was transferred to a glass tube containing and surrounded by liquid nitrogen. This was connected to the addition arm and the liquid nitrogen inside the tube pumped off, while the temperature of the tube and sample was kept at 78°K by the liquid nitrogen surrounding the tube.

The calorimetric procedure from this point on and the calculations were, with minor modifications, those described previously.[5] The modifications in the equipment, procedure, and calculations will be reported in detail elsewhere.

Tensile Tests

The tensile tests were made on samples of the alloy rolled into strip of suitable thickness. After shaping in a jig, marking with gauge marks and labelling, all samples were annealed at 600°C for 30 minutes to obtain a consistent grain size.

The stress on the sample was exerted by a dead load and measured by a calibrated strain ring and recorder assembly. The extension corresponding to a given stress was measured with a cathetometer; the length after fracture was determined by direct measurement of the specimen. The tensile testing equipment, which was built for another investigation,[6] was designed in such a way that the test piece and adjacent parts could be immersed in a liquid in order to maintain a constant temperature. In the present application the tests were made either at room temperature in air or at 78°K under liquid nitrogen.

TABLE II. Calorimetric measurements of stored energy.

Run No.	Temperature	Type of drill	cal/g
40	R.T.	$\frac{33}{64}$ twist drill	0.53
41*a*	R.T.	$\frac{3}{8}$ S.F. drill	0.50
87	78°K	$\frac{3}{8}$ S.F. drill	1.47
92	78°K	$\frac{3}{8}$ S.F. drill	1.51
94	78°K	$\frac{3}{8}$ S.F. drill	1.47
96	78°K	$\frac{33}{32}$ twist drill	1.35

[6] E. J. Suoninen, "Investigation of the martensitic transformation in metastable beta brass," S. M. thesis, Dept. of Metallurgy, M.I.T., 1954.

EXPERIMENTAL RESULTS

Cold Working

The energy expended in cutting the alloy is given in Table I. The figures reported represent the average of several tests and have a mean deviation of about 10 percent. They were calculated from the measured values of the torque.

Calorimetry

Table II shows the results of the calorimetric measurements on the deformed chips. The results of preliminary runs, made at both temperatures of cutting, were considered less dependable and are not reported here.

Tensile Tests

The results of the tensile tests up to the point of maximum load are given in Fig. 2 as a plot of true stress against true strain for one specimen of the alloy deformed in tension at room temperature and two specimens deformed at 78°K. Other tests showed a small scatter on either side of those reported here. The curves are carried only to maximum load, since necking prevented measuring the true stress beyond this point up to fracture. The elongation at fracture, as measured directly, was 30 percent after extension at room temperature and the corresponding value for 78°K was 42 percent.

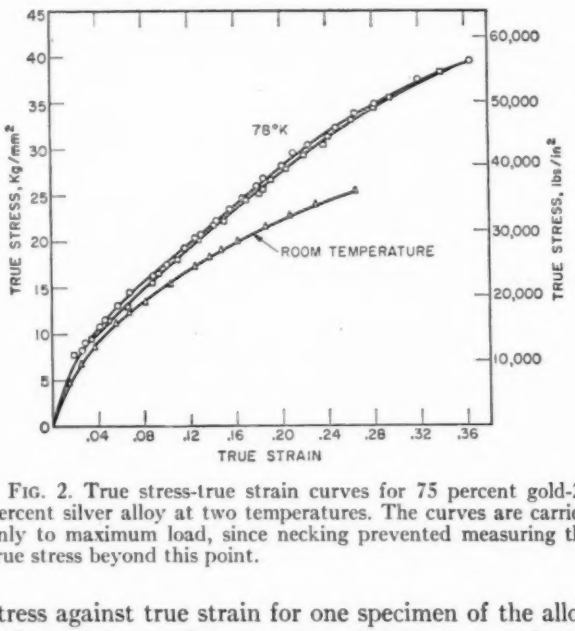


FIG. 2. True stress-true strain curves for 75 percent gold-25 percent silver alloy at two temperatures. The curves are carried only to maximum load, since necking prevented measuring the true stress beyond this point.

DISCUSSION OF RESULTS

Cold Working

From previous results[3] it is reasonable to assume that not more than 10 percent of the total energy is

required to overcome friction and the remainder is available for deforming the alloy. It is also characteristic of this soft gold-silver alloy that only high rake angle tools produce thin, smooth chips. The rake angles used were only about 15° for twist drills and 0° for straight-fluted drills. The chips, therefore, were too irregular to measure, but it was evident that they had been deformed severely. This accounts for drilling energies at room temperature of the same order as for mild steel, even though steel has twice the tensile strength of the alloy used in these experiments. From an engineering standpoint it should be realized that the results obtained with the alloy cannot be carried over directly to other metals.

Stored Energy

The energy stored by the alloy due to drilling at room temperature is larger than the values measured for other deformation processes carried out on this alloy. The value of 0.52 cal/g for drilling should be compared with the measured maximum values of approximately 0.17 cal/g for rolling strip[2] and 0.34 cal/g for orthogonal cutting.[3] The difference in the stored energies due to orthogonal cutting with a 15° rake angle and drilling with a twist drill having the same rake angle may be attributed, at least in part, to the difference in strain rate or the difference in the method of cooling of the chips or both.

The most important finding of this investigation is that the energy stored due to drilling at 78°K is roughly three times the value obtained at room temperature. The lower value of two and one-half times for the twist drill depends on one measurement only (due to the experimental difficulties in handling the chips produced by this drill at 78°K) and less significance is, therefore, assigned to it.

The literature does not seem to record any other investigation which permits a comparison of the energy of cold work stored at room temperature and at a lower temperature under otherwise identical conditions. In one investigation,[7] copper was deformed at 193°K and at 93°K, but the measurements of the stored energy were started at room temperature. The results showed that recovery and attendant evolution of stored energy had started and in some cases had finished below room temperature and the total stored energy, therefore, could not be measured. Another investigation[8] measured the energy stored during compression at the temperature of liquid nitrogen. The values were 1.13 cal/g for cadmium and 0.53 cal/g for lead, but no data are given for the energy stored in these metals during working at room temperature.

[7] H. Kanzaki, J. Phys. Soc. Japan **6**, 456 (1951).

[8] Khotkevich, Chaikovskii, and Zashkvara, Doklady Akad. Nauk S.S.S.R. **96**, 483 (1954).

Stress-Strain Curves

The true stress-true strain curve of the alloy for room temperature follows a normal pattern for nonferrous metals in that it never becomes a straight line. The degree of strain hardening is greater in tensile tests at 78°K than at room temperature. The same is true of the tensile strength and the elongation at fracture. The toughness, as measured by the area under the stress-strain curve, doubles with the decrease in temperature. These results confirm a generalization for face-centered cubic metals, according to which a decrease in temperature is accompanied by an increase in strength without a decrease in ductility.[9,10]

General Comments

As can be seen from Tables I and II, the energy of deformation increased by a factor of approximately two with the decrease in temperature, while the stored energy increased by a significantly larger factor. This increase in stored energy, therefore, is not attributable solely to the increase in the energy of deformation. This is also shown by the essentially identical amounts of energy stored at room temperature, even though the energies required for the deformation differed by a factor of two for the two types of drill.

The twofold increase in the energy of deformation resulting from the change from room temperature to 78°K may be compared with an increase by a factor of only 1.4 in the true stress obtained by extrapolating the tensile curves to the strains expected in cutting. Obviously the two types of deformation are not equivalent.

The fundamental differences between cold worked states which originate by working at different temperatures, such as room temperature and 78°K, may be due to several causes. In particular, they may be caused by differences in the type of imperfections or in their number and distribution. The increase in stored energy with decreasing temperature found in this investigation can be explained by either or both of these differences. While the values of the stored energy do not by themselves provide a basis for distinguishing between these alternatives, they are part of the body of observations for which any theory of cold working must account.

ACKNOWLEDGMENTS

The authors wish to record their indebtedness to Mr. L. R. Rubin and Dr. Peter Greenfield for advice and assistance. Thanks are also due Mr. Benjamin Howell for his contributions to the experimental work. The research was performed under Contract No. AT-30-Gen-1002 with the U. S. Atomic Energy Commission.

[9] W. J. De Haas and R. Hadfield, Phil. Trans. Roy. Soc. London (A) **232**, 297 (1933).

[10] P. L. Teed, *The Properties of Metallic Materials at Low Temperatures* (John Wiley and Sons, Inc., New York, 1950).

Work Function of Tungsten Single Crystal Planes Measured by the Field Emission Microscope*

ERWIN W. MÜLLER
Physics Department, The Pennsylvania State Univeristy, University Park, Pennsylvania
(Received October 4, 1954)

The work functions of the low emitting crystallographic planes of tungsten, (011) and (112), are not well known. In previous experiments scattered secondaries and some other disturbing effects covered the small emission particularly of the (011) plane. A method combining the field emission microscope with a probe collector has been applied. The current density in the (011) plane turned out to be 4 to 5 orders of magnitude smaller than in strongly emitting planes. Applying the Fowler-Nordheim theory, ϕ_{011} was found between 5.70 and 5.99 ev and ϕ_{112} between 4.65 and 4.88 ev depending on the temperature at which the tungsten crystal has been annealed previously. This is not a temperature effect of the work function, but the result of freezing in thermal imperfections of the planes. The work function of the ideal (011) plane appears to be as high as 5.99 ev. The existence of such a high value is further suggested by the ionization of aluminum on a hot polycrystalline tungsten wire, indicating that a part of the surface has a work function as high as the 5.96 ev ionization energy of aluminum.

INTRODUCTION

ONE cannot say that our knowledge of electronic work functions of metals is well established if one considers the fact of the crystallographic orientation of the surface. Actually, the only reliable experiments on the work function of different crystal planes have been made with tungsten. However, though this metal may be outgassed thoroughly, the data obtained recently vary considerably, e.g., between 4.65 and 5.5 volts for (011) as the basic plane of the lattice. It may be recalled that the average effective work function of a polycrystalline tungsten surface is well agreed upon to be 4.54 volts with an error of probably only 0.02 volt. Also, there is good agreement among various authors for the strongly emitting planes. The experimental difficulty is to measure the much smaller emission of the planes with a higher work function, particularly (011) and (112), and perhaps (100).

The application of field emission to our problem offers some advantages. There is good reason[1,2] to assume the validity of the Fowler-Nordheim[3] theory which suggests plotting $\log j/F^2$ *versus* $1/F$ and gives then the $\frac{3}{2}$ power of the work function as the slope. The development of the smoothly pointed field emitter[4] and the field emission microscope[5] provided a cathode with the following advantages of special importance for our problem:

(1) All crystallographic planes are accessible on one cathode.

(2) A clean surface can be obtained easily and can be observed at the instant of the measurement.

(3) The work function is measured at room temperature, so that a possible temperature dependence is not encountered as in the case of evaluating thermionic emission by the Richardson plot.

(4) The emission is obtained from a nearly perfect surface area of about 100 A diameter, with a minimum of undesired macrostructure such as shingles, pits, and patches.

The main difficulty of field emission is the determination of the true field strength at the cathode and the true current density. With the present means it is not yet possible to measure absolute work functions accurately enough. However, since the work functions of the strongly emitting planes (116), (130), and (111) are already well known, we can make relative determinations for all other planes using the advantages of the field emission method.

EXPERIMENTAL METHOD

In the first communication about the field emission microscope[5] the nonuniform distribution of the current density J over the surface of the hemispherical tungsten monocrystal had already been attributed to the variation of the work function over the different planes. A quantitative measurement had then been made by measuring the current going through a probe hole in the screen.[6,7] This seems to remain the only reliable method to deal with the problem of measuring current densities in the emission pattern. Some recent investigations[8,9] have apparently overlooked the warnings that have been expressed by the author[6,10] regarding direct measurements of screen light output for this purpose. The application of this latter method is very tempting indeed because of its simplicity. However, in spite of the complication which arises from the necessity

* This research was supported by the U. S. Air Force, through the Office of Scientific Research of the Air Research and Development Command.

[1] R. Haefer, Z. Physik **116**, 604 (1940).
[2] W. P. Dyke and J. K. Trolan, Phys. Rev. **89**, 799 (1953).
[3] R. H. Fowler and L. W. Nordheim, Proc. Roy. Soc. (London) **A119**, 173 (1928).
[4] E. W. Müller, Z. Physik **106**, 132 (1937).
[5] E. W. Müller, Z. Physik **106**, 541 (1937).
[6] E. W. Müller, Z. Physik **120**, 261 (1943).
[7] M. Drechsler and E. W. Müller, Z. Physik **134**, 208 (1953).
[8] M. K. Wilkinson, J. Appl. Phys. **24**, 1203 (1953).
[9] Dyke, Trolan, Dolan, and Grundhauser, J. Appl. Phys. **25**, 106 (1954).
[10] E. W. Müller, Ergeb. exakt. Naturw. **27**, 327 (1953).

of a calibration of the screen response for various electron energies, one cannot be at all sure that the light output from an area of the screen is a simple function of the electron density in the beam originally emitted by the cathode towards this area. In fact, the dark areas on the pattern are brightened by scattered light from the bright areas of the picture, both by reflection from the inner walls of the tube and inside the phosphor, and further by total reflection inside the glass wall. Another very disturbing brightening of the dark areas is due to secondary electrons released somewhere at the screen or at the anode. Also the excitation of the dark screen areas by soft x-rays within the tube cannot be neglected. It is therefore not astonishing that Wilkinson[8] found only a contrast of $J_{(013)}/J_{(011)}=30$, corresponding to a work function of 4.58 volts for (011). We shall see later that the contrast actually amounts to 10^4 to 10^5. Dyke *et al.*[9] applied some measures to eliminate the optical scattering effects by backing the screen with aluminum and by coating the outside wall of the tube to avoid total reflection. This raised the contrast $J_{(013)}/J_{(011)}$ immediately to 1600, corresponding to $\phi_{(011)}=5.2$ volts. The protection against optical scattering does not reduce the disturbing effect of secondaries and x-rays.

Considering these circumstances the old method, with a probe hole in the screen and direct measurement of the electron current, has been resumed. Two such tubes are shown in Fig. 1(a) and 1(b). In order to bring the different parts of the picture onto the probe hole, either the anode assembly with the hole is made mobile as in Fig. 1(a) or the tip can be rotated as in Fig. 1(b). The cathode can be replaced more easily in design in Fig. 1(a), but the anode assembly can be shifted only to a limited extent. Also the electron beam through the hole always hits different parts of the collector. The design in Fig. 1(b) makes it more difficult to replace a cathode tip and only permits covering a stripe of the pattern. The advantages are that the tip can be turned to larger angles, and the beam always hits the same spot on the collector.

The collector, which has to pick up currents down to 10^{-15} amp, is highly insulated and is close to ground potential (+40 volts), the seal being surrounded by a guard ring at the same potential. The collector is completely shielded by the suppressor being held at a potential of +20 volts with respect to the cathode. This eliminates secondaries possibly released somewhere at the anode. The suppressor also allows smooth operation with the FP54-electrometer amplifier, acting as a screen grid against charges induced by changes of the anode high voltage potential during the measurements. In the design according to Fig. 1(a), the glass wall in the area opposite to the gap between the anode and the suppressor was covered with a conductive coating of evaporated molybdenum, being held at suppressor potential. This was not necessary in the more closed design of Fig. 1(b).

The hole in the screen has a diameter from 2.5 to 1.5 mm in the different tubes. The phosphor coating covers the bottom of the anode cup, leaving free a margin of about 1 mm around the probe hole. This

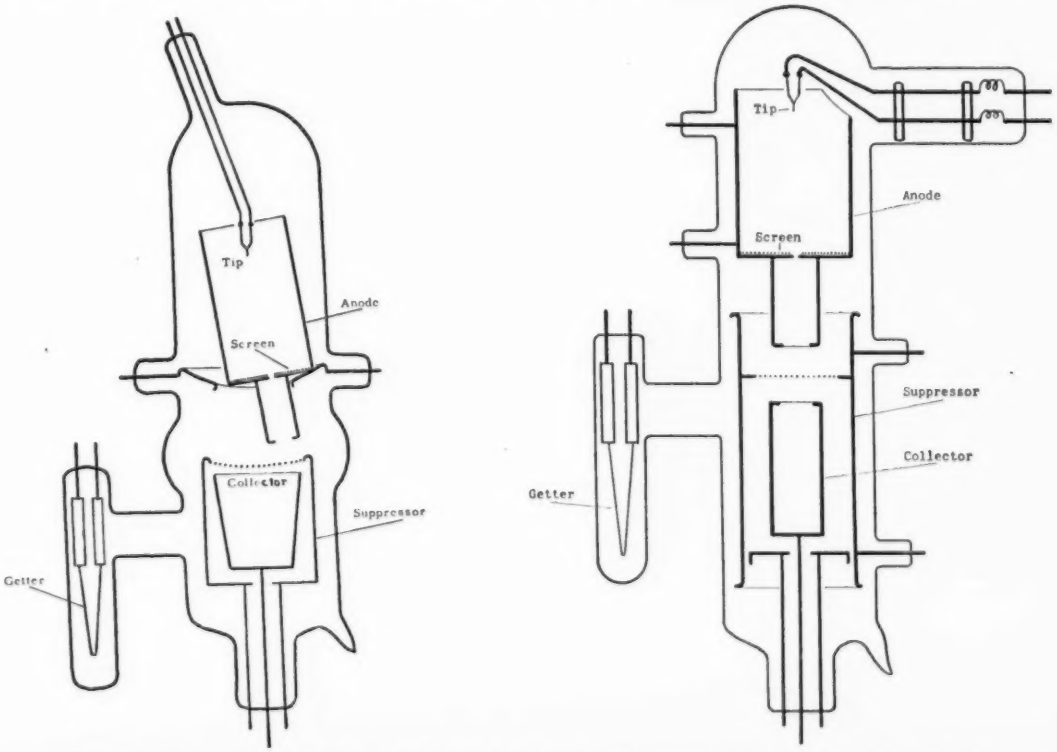


FIG. 1. Experimental field emission tubes: (a) with adjustable anode, (b) with adjustable cathode.

FIG. 2. Tungsten pattern, (011) in center, with probe hole adjusted to (012) plane.

prevents a lens effect on the passing electron beam because of the voltage difference across the screen thickness. The crystallographic plane to be measured is adjusted onto the probe hole by shifting either the anode assembly or the cathode, while the fine adjustment to the exact center of the plane can be made by deflecting the pattern magnetically. Figure 2 gives an indication of the size of the probe hole relative to the apparent size of the crystallographic planes. The penumbra around the dark hole comes from the light reflected by the uncoated nickel sheet carrying the screen.

All metal parts of the tube are thoroughly outgassed before assembly. The tube is then baked out for several days. The metal parts are heated by rf to red heat, including the screen, which withstands this treatment if the phosphor is zinc silicate or calcium tungstate. The cathode and the getter filaments are heated close to 2700°C. The baking out procedure is then repeated for several hours, and after another heat treatment of the filaments the tube is sealed off. The evaporation of a part of the tantalum getter filaments ends the evacuation process, yielding a sufficient vacuum to keep the field emission voltage-current characteristic unchanged during the time required for the measurements. For a given current the anode voltage changed by 0.1 percent in 24 hours (without heating the cathode in between, and with a current of 1 μA drawn for only 1 min each time). No adsorption film was visible on the pattern. This indicates a vacuum better than 10^{-12} mm of gases which can be adsorbed on tungsten at room temperature.

MEASUREMENTS

The first task was to determine favorable voltages for the suppressor and the collector. Current-voltage characteristics for different suppressor voltages were taken. It was found that with a collector potential of 40 volts the suppressor voltage may lie between 10 and 30 volts without varying the collector current appreciably. The measurements were finally taken at a suppressor potential of 20 volts while the collector was kept at 40 volts. The voltage drop across the input resistance of the electrometer amplifier was always less than 0.05 volts and thus negligible.

Having adjusted the desired crystallographic plane on the anode hole, the procedure for measuring the characteristics began by annealing the tip to clean it once more. After the cathode was cooled down, total emission currents between 10^{-5} and 10^{-8} amp were drawn and the anode voltage as well as the collector current were read. Since it has been found that the annealing temperature influenced the characteristics for the (011) and (112) planes, this temperature was varied in different readings. In each of the four tubes used for this work, the emission of all important crystallographic planes that were accessible were measured. The only missing plane is (100) which none of the tubes allowed to be centered on the hole. All our tips have the (011) direction on the axis, and the (100) plane is 45° off, too far to be reached with the design Fig. 1(a). None of the patterns obtained with the tube in Fig. 1(b) had an azimuth favorable for reading (100), although the tip was replaced three times. For the determination of a possible temperature dependence of the work function, several runs were made with the tip at elevated temperatures. However, one could not go higher than 500°C because of the beginning deformation of the tip under the influence of surface migration and electrostatic field forces.

ANALYSIS OF CHARACTERISTICS

The wave mechanical theory[11] gives the following equation for the current density J in amp/cm² as a function of the field strength F in volts/cm and the work function in ev:

$$J=\frac{1.55\cdot 10^{-6}F^2}{\phi}\cdot\exp\left(-\frac{6.86\cdot 10^7\phi^{\frac{3}{2}}}{F}\right)\cdot v(y) \qquad (1)$$

where $v(y)$ is Nordheim's elliptic function of the variable $y=(3.62\cdot 10^{-4}F^{\frac{1}{2}})/\Phi$ in the form as corrected by Burgess, Kroemer, and Houston.[12] It is convenient to plot the measured field emission characteristics as $\log J/F^2$ *versus* $1/F$, which gives almost straight lines since $v(y)$ varies only by a small amount in the accessible narrow range of field strength. The slope of the Fowler-Nordheim characteristic is:

$$\alpha=\frac{d\log_{10}J/F^2}{d(1/F)}=-2.98\cdot 10^7\phi^{\frac{3}{2}}\left(v-\frac{y}{2}\frac{dv}{dy}\right) \qquad (2)$$

$$\alpha=-2.98\cdot 10^7\phi^{\frac{3}{2}}s(y). \qquad (3)$$

[11] A. Sommerfeld and H. Bethe, Handbuch d. Physik **24**, 2, 438 (1933).

[12] Burgess, Kroemer, and Houston, Phys. Rev. **90**, 515 (1953).

The function $s(y)$ has been tabulated by Burgess, Kroemer, and Houston. It varies only between 0.95 and 0.98 for the range of field strength and work function that is being covered with our measurements. This allows us to convert the measured voltages and total currents into field strengths and current densities as required for the Fowler-Nordheim plot by successive approximation. We then relate the data of the probe current to the characteristic for the total current, for which an average effective work function of $\phi=4.50$ ev and an emitting area $2r^2$ is assumed. The field factor $B=F/V$ is then adjusted by about 3 to 5 percent to give the total current plot the theoretical slope $a=-2.98\cdot10^7\cdot4.50^{\frac{3}{2}}s(y)$. With this factor B the Fowler-Nordheim plot for the single crystallographic planes can be drawn from the experimental current-voltage characteristics. The slope of this almost straight line gives then the desired work function according to Eq. (3).

For the first approximation described above, the radius of the tip is calculated with a semiempirical formula by Drechsler and Henkel,[13] $r=0.0074V^{\frac{1}{2}}$, which gives r in A units if V is the voltage required for 10^{-5} amp emission. A plot of the theoretical values of $\log J$ *versus* F, which have been handily tabulated by Dolan,[14] is then used to find the approximate field strength producing the average current density $10^{-5}/2r^2$ amp/cm^2. The ratio $B=F/V$ allows then the first Fowler-Nordheim plot to be drawn for the total current.

The current density in the area cut out by the probe hole is calculated by dividing the collector current by the size of the hole projected back onto the tip surface. This requires knowing the magnification of the pattern, which can be measured by determining the distance from the (112) to the $(\bar{1}21)$ plane on the screen which equals r times the magnification.

Since we compare only the relative slopes of the Fowler-Nordheim characteristics, an error in the size of the tip radius or in the effective size of the probe area is quite negligible. Also, the measurement of the total emission as a reference is merely a matter of convenience, one could as well compare directly the collector currents for various crystal planes. The prerequisite of a uniform field strength across the whole tip, however, is not completely fulfilled. There are two different details to be considered.

(1) The geometric shape of the tip, representing something between a hyperboloid and a spherical bulb on a cone[13] produces a drop of the field strength with increasing angle from the apex. The degree of the drop depends on the special shape of the tip. In the present work only "wide angle" tips have been used, that is tips having a neck behind the spherical end. They show at the periphery of the pattern the (011) and the $(01\bar{1})$ planes simultaneously, which are opposite each other, 180° apart. Such tips can be made by a proper etching process. They give a very slow drop of F with increasing apex angle. The drop can be determined quite accurately by comparing the slopes of pairs of planes with the same Miller indices, such as $(\bar{2}33)$ and $(\bar{3}32)$, and different angular distances θ from the apex. A drop of 0.15 percent per degree was found in good agreement with Dyke and Trolan.[2] This figure was used for the conversion of the measured voltages into true field strengths on the different planes.

(2) Electron microscope studies in which the tip could be turned around its axis revealed that there is a slight deformation of the hemispherical cap at the location of the (011) planes. Field emission microscopy shows that these planes are completely flat within an angle of $\theta=5.5°$ to 7° from the axis. The resulting reduction of the field strength at the flat center of the spherical cap was determined by two ways. A metal model of a tip was made consisting of a rod 3 in. diameter ending in a hemisphere. The anode was represented by a metal sheet cap of 7 in. radius of curvature concentric to the tip hemisphere. The field strength which appeared at a given voltage between the tip and the anode was determined by measuring the mechanical force $P=(F^2/8\pi)$ dyne/cm^2 (F in electrostatic units). This could be done by observing the voltage which was just sufficient to lift up a $\frac{1}{4}$ in. diameter disk of aluminum foil placed on the apex. This method should yield accurate values because of the dependence on the square of F. The adhesion forces of the aluminum foil to the metal tip were made less erratic by coating the tip model with graphite. Fifty measurements were taken, therefore, and a distribution curve plotted which had a width of about 600 volts at a total voltage of 9100 volts. A second and a third series of measurements were then taken with the tip hemisphere flattened to an angle of $\theta=7°$ and 10°, respectively. The new distribution curves were very similar to the first one except for a shift of 750 or 1300 volts, respectively, representing a drop of the field strength by 6 or 14 percent caused by the flat on the apex. Plotting this as a function of the half-angle θ of the flat we find for our (011) planes with $\theta=7°$ a field drop of 6 percent in the apex, and for $\theta=5.5°$, 3 percent. The latter value agrees with the result of another method. Under certain conditions, an absorption film of aluminum oxide on tungsten shows in the field emission microscope a homogeneous layer on the (011) plane with roughly hexagonal periphery and without any details in it except a gentle drop of intensity towards the center. The current density of the emission of this film corresponds to about $\phi=4.5$ ev, and the dip in the center owing to the field drop was measured photometrically to be 0.65 at $40\cdot10^6$ volt/cm and 0.50 at $30\cdot10^6$ volt/cm. Both values correspond to a field drop of 3 percent in the center of the (011) plane. This figure was then used in the final reduction of the (011) characteristic.

[13] M. Drechsler and E. Henkel, Z. angew. Phys. 6, 341 (1954).
[14] W. W. Dolan, Phys. Rev. 91, 510 (1953).

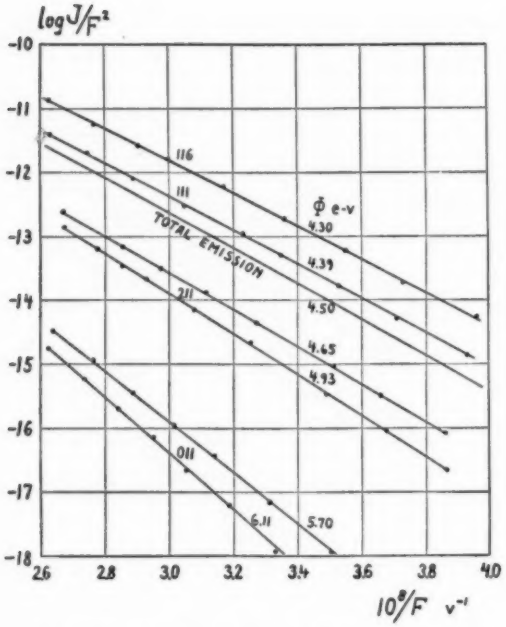


FIG. 3. Fowler-Nordheim plots of field emission from various planes.

RESULTS

Fowler-Nordheim plots of some characteristics are presented in Fig. 3. Field reductions of 3 percent for the (011) plane with 5.5° half-angle, and of 0.5 percent for the (112) plane with 2.5° half-angle, and for the various distances from the apex are taken into account. Table I presents the data obtained with a tube according to Fig. 1(b). These are averages from 3 to 5 runs at different tip radii. The results of the other tubes fall within the range of the error for a single measurement which is estimated to be less than 1 percent for the strongly emitting planes and 1.5 percent for the (011) plane. The recent results of thermionic measurements by Smith[15] are given for comparison. The agreement at the strongly emitting planes (116) and (111) is perfect, so that our assumption of an average work function of 4.50 volt seems to be justified. A striking difference, however, appears at the (112) and particularly at the (011) plane. The high value for the latter was not unexpected, since 5.7 to 6.0 ev was found with the same

TABLE I.

Plane	ϕ Field emission	ϕ Smith
(116)	4.30	4.29
(013)	4.31	...
(012)	4.34	...
(122)	4.35	...
(111)	4.39	4.39
(233)	4.46	...
(123)	4.52	...
(112)	4.65 to 4.88	4.65
(011)	5.70 to 5.99	5.26 estimated

[15] G. F. Smith, Phys. Rev. **94**, 295 (1954).

method, though less elaborate, several years ago by the author.[6,7] It is clear that other recent attempts to measure $\phi_{(011)}$ must have been inaccurate because of the failure to avoid secondary emission, light scattering, and x-rays which cover the weak emission of that plane. Less comforting was the erratic scattering of the data until it was found that the value of ϕ depended on the annealing temperature which had been applied to the tip before the run. Figure 4 shows a graph of the data for $\phi_{(011)}$ and $\phi_{(112)}$ as a function of the annealing temperature. The tip was heated for 2 to 5 seconds in the high temperature range and up to 10 minutes in the lower range, and then the heater current suddenly turned off. Annealing a tungsten tip at low temperatures increases the size of the low emitting planes.[16] Our measurements with a field ion microscope,[10,17] which shows directly the edges of the net planes, give an increase of the half-angle θ of the flat from 5.5° to 7° at the (011) plane, and of 2.5° to 4.5° at the (112)

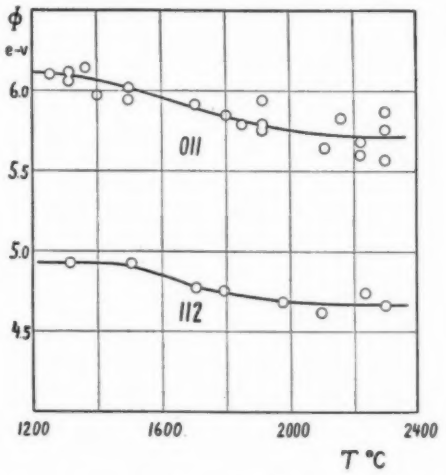


FIG. 4. Work function of (011) and (112) plane as a function of the temperature at which the cathode was annealed before the measurement.

plane, when the annealing temperature is lowered from 2200°C to 1200°C. This reduces the field in the center of the (011) plane additionally[18] by 3 percent, and of the (112) plane additionally by 1.5 percent. In the plot, Fig. 4, the work function of (011), as frozen in from an annealing temperature of 1200°C, has then to be reduced by 2 percent to 5.99 ev, and the work function of (112) by 1 percent to 4.88 ev. However, there remains still a considerable dependence on annealing temperature to be attributed to other causes. At the high temperatures the surface is apparently in a dynamic state of considerable disorder, and the random distribution of missing or protruding atoms in and on the first surface layer is frozen in when the temperature is suddenly lowered. This "rough" sur-

[16] J. A. Becker, Bell System Tech. J. **30**, 907 (1951).
[17] E. W. Müller, Z. Physik **131**, 136 (1951).
[18] The author is indebted to Dr. J. M. Houston for this suggestion.

face has a relatively low work function as do the other loosely packed planes. Annealing at lower temperatures heals the imperfections of the closely packed (011) plane and gives the true work function of about 6.0 ev. We cannot yet tell how fast the healing process approaches the low temperature equilibrium state after turning off the heater current. About one-tenth of a second is required to cool the tip below the temperature of considerable surface migration, about 1000°C. It seems not impossible that at the temperature at which thermionic emission has been measured, for instance by Smith, the effective work function of the (011) plane owing to the thermal imperfections is really as low as his estimate of 5.26 ev indicates.

In the case of field emitters the surface mobility is probably larger than on the surface of a thermionic cathode. This is due to the drop of the surface energy with increasing distance from the apex because of the steep change of the curvature. If a tip is heated the edges of the net planes around (011) are continuously dissolved and the atoms migrate fast to the rear of the cone.[19] This effect has been demonstrated with the field emission microscope by absorbing a certain amount of SrO on the tip, which marks particularly the edges of the (011) planes.[10] If the tip is heated, these dissolving lattice steps move in concentric rings towards the center of (011). The moving lattice steps may just stop across the probe hole if the heat is turned off and thus cause the erratic data in Fig. 4. Without an activation these lattice steps are invisible in a normal field emission, microscope with its insufficient ability to overcome the large contrast. To prove their presence, a tube was built according to Fig. 5 using the same principle as the tubes of Fig. 1, with the only difference that the collector was replaced by a fluorescent screen on a higher potential and that the probe hole was 4 mm in diameter, covering a larger part of the (011) plane. The anode aperture appeared enlarged on the second screen, and if the tip was heated to about 1300°C, the dissolving rings could be seen disappearing at a rate of about one in five seconds. This tube could be used for measuring the relative current density in (011) and (012) for comparison. With a suppressor potential of -100 volts with respect to cathode, the contrast in the two crystallographic directions was 1:15 000 as measured by a photomultiplier at 5 μA total emission current. However, there was still too much reflected light from the very bright first screen coming through the anode aperture, so the actual contrast was probably not smaller than that measured electrically in the tubes shown in Fig. 1.

With the surprisingly high work function found for the (011) plane of tungsten, the question arises as to whether one could check this result by any other method. It would be very desirable to make contact potential measurements on tungsten single crystals.

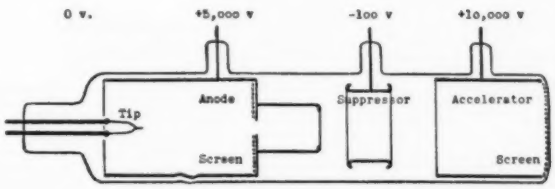


FIG. 5. Field emission microscope for observation of the (011) plane.

However, it is not sure to what extent the results would be spoiled by the inevitable microscopic defects of the surface. Another possibility is the investigation of the surface ionization effect. A preliminary experiment has been made by re-evaporating aluminum from a hot polycrystalline tungsten coil. Roughly 10 percent of the evaporated aluminum atoms left the surface as positive ions. Since the ionization potential of Al is 5.96 volts, we may assume that a part of the polycrystalline surface had a work function as high as this value. When in the same tube the aluminum was replaced by copper with an ionization potential of 7.65 volts, no surface ionization was observed. Of course there was no oxide film on the tungsten surface.

DISCUSSION

Tungsten is certainly not the only metal that presents differences of the work functions of various crystallographic planes up to 1.7 ev. The field emission microscope patterns of many other metals appear with similar contrasts, the closely packed plane having always the highest work function. It is astonishing that these large differences were not recognized earlier. Smoluchowski[20] has given a detailed theoretical calculation of the work function of tungsten crystal planes. By assuming not more than $\frac{1}{6}$ free electron per atom, he found an agreement with Nichols'[21] experimental data, which were the only available ones at the time. With an assumed electron density of 0.5 to 0.8 electrons per atom the data of this theory could be roughly matched with our experimental results, but the ratio of $\phi_{(111)}$ and $\phi_{(112)}$ does not fit too well, probably because the simple model of the theory did not take into account the complicated effect of the d electrons. Their contribution is clearly established by the comparison of the field emission patterns of W and Mo with those of Ta and Nb. All four metals show the characteristic pattern of the body centered cubic lattice. W and Mo have exactly identical patterns. Ta and Nb form another group with (112) planes emitting almost as strong as (111). Obviously the difference in the d–s electron configuration is responsible for the change.

ACKNOWLEDGMENTS

Thanks are due Miss Elizabeth Hebelka and Mr. Walter Pimbley for assistance in performing measurements.

[19] E. W. Müller, Z. Physik **126**, 642 (1949).

[20] R. Smoluchowski, Phys. Rev. **60**, 661 (1941).

[21] M. H. Nichols, Phys. Rev. **57**, 297 (1940).

Viscoelasticity of Poly Methacrylates

F. BUECHE*
Rohm and Haas Company, Philadelphia, Pennsylvania
(Received October 20, 1954)

This paper presents an intensive study of the viscous and elastic properties of the poly methacrylates. Variation of these properties is studied as a function of temperature in the range $60<T<160°C$, as a function of molecular weight in the range $5\times10^3<M<5\times10^6$, and as a function of plasticizer concentration over the range 0–100 percent. Capillary viscometers were used in the low-viscosity systems and tensile creep and recovery measurements under very small loads were used to study the less fluid systems. Principal effort was confined to polymethyl methacrylate but enough work was done using the butyl and decyl esters to show that the whole series of these polymers fits into a relatively simple picture. Consequently, given a certain minimum amount of data, it is now possible to predict the viscoelastic behavior of these polymers in the range of variables given above.

The theory of viscoelasticity previously given has been extended so as to take account of the effects of chain entanglements and distributions in molecular weight. Moderately good agreement is found between theory and experiment.

INTRODUCTION

UNTIL a relatively short time ago, the molecular mechanisms responsible for the viscoelastic properties of plastics were only poorly understood. Consequently, experimenters in this field were concerned mainly with obtaining a broad survey of the phenomena involved. With the advent of a more firm theoretical picture of the molecular mechanics underlying the process of polymeric viscoelasticity,[1,2] we are now in considerable need of a more intensive experimental investigation.

We have set out in this work to provide a comprehensive picture of the viscoelastic behavior of polymethyl methacrylate and, to a lesser extent, the higher methacrylates. It is necessary at this time to restrict ourselves to behavior under small to moderate loadings and to temperatures not too far below the glass temperature, T_g, of these polymers.

In this study we have examined the viscoelastic behavior of several poly methacrylates in quite some detail. Care has been taken to use fractionated materials whenever it seemed advisable so as to decrease the complexity of the phenomena being observed. In a further effort to make our experimental results easily understandable, we have employed tensile creep methods to obtain the dynamic data. Even though this experimental technique sacrifices the availability of a wide variation of time scale (the convenient range being three decades), we believe the clarity of the results so obtained compensates for this shortcoming.

At the outset of these experiments it became fairly obvious that existing theories of viscoelasticity were inadequate to explain the short time response of high molecular-weight polymers. From the nature of our tensile creep curves we were able to conclude that the effect of chain entanglements was largely responsible for this disparity between theory and experiment. We have extended our previous theory so as to consider this effect and this development, as well as others, are included in appendices to this paper.

EXPERIMENTAL

The polymethacrylates used in this investigation were prepared by bulk polymerization (usually at 60°C) using appropriate amounts of catalyst (azo-bis-isobutyronitrile) and *n*-butyl mercaptan to obtain the desired molecular weight. All polymers were dissolved and reprecipitated at least once to remove monomer. Fractionated samples were obtained by the usual method which involves the selective precipitation of the polymer by adding nonsolvent. In general, the initial solution used for fractionation consisted of a 1 percent or less solution of the polymer in acetone. Either petroleum ether or methanol was used as the precipitant. Since the first fraction might possibly contain a small amount of branched material, it was not used for the measurements reported here.

Molecular weights of the polymethyl methacrylates (PMMA) were obtained from intrinsic viscosity measurements in benzene using the relation[3]

$$\log M=\{\log[\eta]+4.245\}/0.76. \quad (1)$$

Viscosities of the plasticicized polymers were determined by two methods. The less-viscous samples were measured using capillary viscometers employing the technique of Fox and Flory which is described elsewhere.[4] Where practical, the viscosities were also determined directly from the tensile creep and recovery curves as described later in the text. The two methods gave viscosities which agreed in the region where comparison could be made. No corrections for rate of shear effects were made since it was found that these effects were small under the conditions of our experiments. Glass

* Present address: Department of Physics, University of Wyoming, Laramie, Wyoming.

[1] Ferry, Landel, and Williams, J. Appl. Phys. **26**, 359 (1955).

[2] F. Bueche, J. Chem. Phys. **22**, 603 (1954).

[3] T. G. Fox, paper given at the September, 1952 American Chemical Society meeting in Atlantic City, New Jersey.

[4] T. G. Fox and P. J. Flory, J. Am. Chem. Soc. **70**, 2384 (1948).

temperatures were obtained by the usual dilatometric methods.[5]

The tensile creep measurements were made using sheets of polymer usually about 10 cm long, 1.5 cm wide, and 0.05 cm thick. These sheets were prepared by laying down a chloroform solution onto a mercury pool and allowing the solvent to evaporate. When the sample had become reasonably hard, it was cut away from the glass container holding the mercury, placed in an air oven and subjected to a heating cycle such that it was raised slowly (several hours) to a temperature of 135°C and held there for about three hours.

Highly plasticized samples were prepared by sealing the polymer and plasticizer in a glass tube which was then heated and rotated until the polymer had dissolved completely. Dilute solution viscosities were taken on the polymer before and after mixing and no degradation was observed. The more solid samples were prepared by mixing the plasticizer with the polymer during the process of preparing the polymer sheet by casting on a mercury surface. Since some plasticizer was always lost during the drying process, the final sample was conditioned by holding it between glass plates at a temperature near 130°C. Concentrations were determined in these cases by dissolving and reprecipitating the polymer.

The tensile creep measurements were made at constant temperature (±0.1°C) and low humidity. This was accomplished by suspending the sample in a glass cylinder (6 cm in diameter, 50 cm tall) which had been sealed off at one end and immersed in a constant-temperature bath. The lower end of the cylinder was filled with Drierite. Since the lower end of the sample was fastened to the bottom of the cylinder by means of a clamp, the appropriate tension was applied to the sample by running a cord from the clamp on the upper end of the sample out of the cylinder and over a pulley so that the proper weights could be suspended from it. As a result of the fact that elongations of the sample never exceeded 10 percent and were usually about 1 percent, no effort was made to compensate for the change in sample cross section. The elongations were measured using a sensitive cathetometer. Before starting measurements on a sample, it was always conditioned for at least a day by holding it in place in the test chamber at a temperature near the highest value to be used in the experiments on that sample.

TABLE I. The variation of $\log_{10}\eta$ for polymethyl methacrylate ($M=63\,000$) as a function of temperature and plasticizer concentration where η is the shear viscosity in poises.

Wt. percent polymer \ Temp. °C	30	40	60	80	100	120	140
8.8	−0.10	−0.26	−0.56	−0.85	−1.09	−1.31	
15.4	0.75	0.57	0.24	−0.06	−0.33	−0.56	
21.3	1.42	1.20	0.83	0.47	0.17	−0.11	
27.4	2.62	2.22	1.65	1.22	0.90	0.63	
31.9	3.35	2.87	2.20	1.66	1.25	0.90	
35.8	4.01	3.50	2.77	2.18	1.70	1.30	
39.6	...	...	3.00	2.37	1.87	1.45	1.05
42.5			3.46	2.66	2.12	1.69	1.32
50.3			4.77	3.67	2.96	2.43	2.00
59.4			6.4	5.37	4.38	3.64	3.03
68.3						4.65	3.78
75.0		11.5	9.03				
90.0		18.4	14.63	12.1			
100.0			24.3	22.2	18.5	11.70	7.72

[5] T. G. Fox and P. J. Flory, J. Appl. Phys. **21**, 581 (1950).

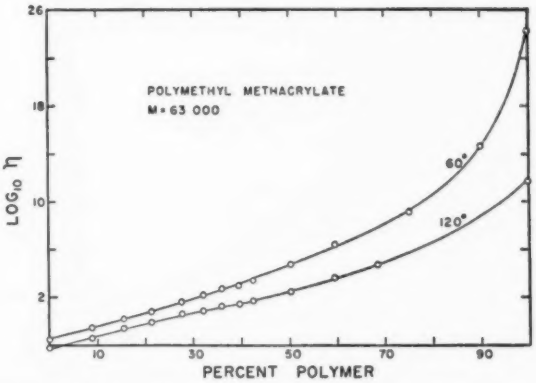


FIG. 1. Variation of the viscosity of polymethyl methacrylate ($M=63\,000$) plasticized with diethyl phthalate at two temperatures, 60° and 120°C.

VISCOSITY

In this section we will examine the following three phenomena: (1) variation of viscosity with plasticizer concentration, (2) dependence of apparent energy of activation for flow on plasticizer content, and (3) variation of viscosity with molecular weight.

To investigate (1) and (2) we chose an unfractionated low molecular weight (63 000) polymethyl methacrylate plasticized with diethyl phthalate (DEP). The polymer had been bulk polymerized at 60°C and was carryied to about 78 percent conversion. This polymer was dissolved and reprecipitated before being used for experiments. Solutions with DEP were prepared as previously outlined.

The viscosities of each solution of polymer in DEP were measured at several temperatures between 30 and 140°C. When these data were plotted as $\log\eta$ *vs* $1/T$ (where η is the viscosity and T is the absolute temperature) it was found that the curves were essentially linear at low-polymer concentrations but acquired considerable curvature as the polymer concentration was increased. The viscosities read from the curves drawn through the experimental points are listed in Table I. In Fig. 1 we have plotted the viscosity as a function of polymer concentration at two temperatures. The viscosities of solutions having polymer concentrations higher than 70 percent were obtained from tensile creep data as described in a later paragraph. All the other viscosities were obtained using capillary viscometers.

The two curves of Fig. 1 should be compared with data previously published for polystyrene solutions

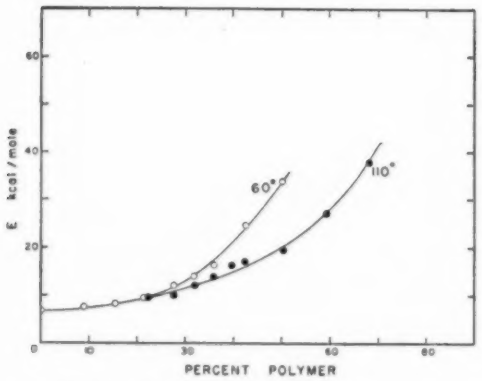


FIG. 2. Variation of the apparent activation energy for viscous flow as a function of diethyl phthalate concentration in polymethyl methacrylate.

in diethylbenzene.[6] The most striking difference between these two materials has to do with the absolute magnitude of the viscosity observed. Comparable polystyrene solutions were found to have viscosities more than 100 times smaller than we observed here for the polymethyl methacrylate solutions. Although some of this difference can be attributed to the difference in solvent, by far the major portion of the difference seems to be a result of the way interchain entanglements act in the two cases. This will become more apparent when one considers that which is to follow.

The apparent energy of activation for viscous flow, E, can be obtained from the slope of the $\log\eta$ *vs* $1/T$ curves. If the curves were linear, E would be independent of temperature. However, since this is not the case at high-polymer concentrations, the value of E will depend upon the temperature at which it is measured. This variation, together with the variation of E with concentration of polymer, is illustrated in Fig. 2. It is observed that, until about 20 percent polymer is added, the value of E is nearly the same as that of the solvent. At higher-polymer concentrations, the activation energy rises rather rapidly and we will see subsequently that it reaches a value near 200 kcal/mole when enough polymer has been added so that the mixture is at its glass temperature provided the temperature is not too low. The factors responsible for this variation in E will be discussed in a later paragraph.

The third point of interest is the variation of η with polymer molecular weight. This variation is illustrated in Fig. 3 where we have plotted $\log\eta$ *vs* $\log M$ for a series of polymer fractions at a fixed polymer concentration of 25 percent by weight.

It is observed that the data at 25 percent polymer concentration may be represented by two intersecting straight lines. The equations of these lines are

$$\begin{aligned} \log\eta_{60} &= 3.4\log M - 14.2 \quad M > 42\,000 \\ \log\eta_{60} &= 1.4\log M - 5.5 \quad M < 42\,000. \end{aligned} \tag{2}$$

[6] F. Bueche, J. Appl. Phys. **24**, 423 (1953).

(We shall indicate logarithms to the base 10 by log and to the base e by ln.) Although these equations hold only at 60°C, the data at other temperatures are parallel to these and exhibit a break at the same molecular weight.

A similar behavior has been found for polystyrene and polyisobutylene.[6,7] It is of considerable interest to notice that the slope in the high molecular-weight region, 3.4, is the same for all three of these polymers. However, the slope of the lower portion is dependent upon polymer type, temperature, and diluent concentration.

The shape of the curve in Fig. 3 has been predicted by theory.[8] It has been shown that the low molecular-weight portion of the curve should be linear with a slope of unity. As such, it represents the normal viscous behavior of "free draining" polymer molecules imbedded in a "solvent" composed of similar polymer molecules together with added plasticizer. Deviations from the expected slope were attributed to the effective loosening of the local liquid structure by the chain ends. For example, a chain segment located somewhere close to a free chain end could move more freely than one where no chain end existed. Indeed, the limited experimental evidence available seems to indicate that, although the slope in this portion is usually found to be larger than unity, it tends to approach unity at high plasticizer content or relatively high temperatures where the effect of chain ends should be less noticeable.

An interpretation of the break in the curve has also been given.[8] It is to be associated with that molecular weight at which each molecule has, on the average, two other molecules entangled with it. For example, the break in our curve occurs at $M = 42\,000$ and so we would infer that the average molecular weight between chain entanglement points in a 25 percent solution of PMMA in DEP is 21 000. (It should be noticed that our notation here is slightly different from that used

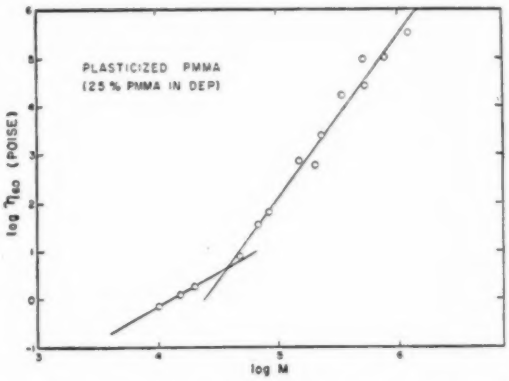


FIG. 3. Variation of bulk viscosity for plasticized PMMA (25 percent PMMA in DEP) as a function of molecular weight. The data were taken at 60°C.

[7] Reference 5 and T. G. Fox and P. J. Flory, J. Phys. & Colloid Chem. **55**, 221 (1951).

[8] F. Bueche, J. Chem. Phys. **20**, 1959 (1952).

in reference 8 where only the entanglements which act to *retard* the motion of the molecule were considered explicitly. Consequently we stated there that M_c was the molecular weight at which, on the average, each chain had one other chain entangled with it—meaning, retarding its motion. In this paper we shall not make such a distinction between retarding and accelerating entanglements.)

The entangled chains will slip somewhat at the points of entanglement and the shape of the upper portion of the $\log\eta$ *vs* $\log M$ curve is determined almost entirely by this slippage factor. If there were no slippage, the viscosity would become nearly infinite at $M=42\,000$. However, the observed slope of the curve indicates that the chains slide over each other in much the same fashion as ropes would do. It is interesting to notice that the same slope obtains for polystyrene, polyisobutylene and PMMA indicating they have the same slippage factor even though the chain side groups are quite different in the three cases. This suggests that the motion of the side groups on the chain is fast enough so that the side group itself is not the point of entanglement.

The data presented in Fig. 3 are for a 25 percent solution of polymethyl methacrylate in DEP. It is of value to know how this curve will change with polymer concentration. From the results obtained with other polymers,[6,9] we can conclude that the slope of the high molecular-weight portion of the curve will remain unchanged until the polymer concentration decreases to about 10–12 percent. The slope of the low molecular-weight portion of the curve should increase as polymer concentration increases since the effect of chain ends will become more pronounced. The critical molecular weight at which the break occurs is inversely proportional to the volume percent of plasticizer present except possibly at very high-plasticizer concentrations. For example, the break in the curve for the pure polymer should occur at $M=(\frac{1}{4})\times 42\,000$ or $M=10\,500$. (We neglect the small difference in density which occurs.) This means that the average molecular weight between chain entanglements is 5250 for pure polymethyl methacrylate. The variation of chain entanglements with polymer concentration may be looked upon as a simple dilution phenomenon. This will become more evident when we consider the effect of diluents and side groups upon the viscoelasticity of methacrylate polymers.

Since the density of pure PMMA is approximately 1.18 under the conditions of our experiment we can summarize the aforementioned statements by the following relation:

$$\eta=KM^{3.4} \text{ for } M>(1.18/\zeta)\times 10\,500, \qquad (3)$$

where K is a constant and ζ is the concentration of PMMA in the solution in gm/cc. In order to find the value of η at various temperatures and molecular weights above the critical entanglement molecular weight, and in the range covered by these experiments, it is only necessary to make use of the data in Table I for $M=63\,000$ and extend the data to other molecular weights by multiplying η by the factor $(M/63\,000)^{3.4}$.

[9] Johnson, Evans, Jordan, and Ferry, J. Colloid Sci. **7**, 498 (1952).

VISCOELASTIC THEORY

When a tensile load is applied to a strip of polymer, four molecular mechanisms contribute to the observed elongation of the polymer sample. The first of these is the ordinary essentially instantaneous deformation associated with the stretching of van der Waals' bonds and bending of valence bonds. This type of deformation is observed in all solids and liquids and is not dependent upon the polymeric nature of our sample. It is nearly temperature independent. In our experiments this type of deformation is extremely small.

A second type of deformation results from the elastic stretching of the network formed in the polymer through the action of interchain entanglements. In the case of pure PMMA, this effect will become important when the molecular weight exceeds the critical molecular weight for chain entanglements, namely 10 500. It is shown in Appendix I that this response is analogous to the response of an actual cross-linked polymeric network.

When the entanglement points begin to move, a third type of elastic deformation becomes important. This is merely a long-range stretching of the polymer chains and would occur even in the absence of entanglements. In general, the maximum elastic extension associated with this type of response is much larger than the network type extension. If a chain has a molecular weight, M, and if the critical molecular weight for entanglements is M_c, then the ratio of the network type extension to the chainlike extension is shown in Appendix I to be $(3M_c/2M)$.

Finally, a fourth type of extension, true viscous flow, must be considered. Whereas the other three mechanisms of extension are reversible (i.e., if the load is removed, the sample contracts and assumes its original dimensions), the extension resulting from viscous flow is nonrecoverable. This part of the deformation is fairly well understood and has been considered in detail elsewhere.[8]

These considerations may be formulated in more precise mathematical terms as shown in the Appendix. When this is done, the relative elongation per unit tensile force on a unit cross-sectional area is found to be

$$(\Delta L/L)/(F/A)$$
$$=(4M_c/3\nu\pi^2MkT)\sum_{p=1}^{s/2}(2p-1)^{-2}[1-\exp(-t/\tau_p)]$$
$$+(32/3\pi^4\nu kT)\sum_{n=1}^{N/2}(2n-1)^{-4}$$
$$\times[1-\exp(-t/\tau_n)]+t/\eta,$$

where

$$\tau_p=4\eta_0 M_c^2/(2p-1)^2\pi^2\nu kTM^2$$
$$\tau_n=4\eta/(2n-1)^2\pi^2\nu kT, \quad (4)$$

k is Boltzmann's constant and n and p are integers.

The other symbols used are: ν, the number of chains per unit volume; σ, the number of chain segments between entanglement points; N, the number of segments in a chain; η, the measured viscosity of the polymer; and η_0, the viscosity which would exist if there were no chain entanglements. To find η_0, one must extrapolate the lower line in Fig. 3 to molecular weights above the break in the curve. Whether one should use the theoretical slope of unity for this purpose or the experimental slope which prevails below the break is not completely clear. In all probability, neither procedure is completely correct since, to be precise, some correction should be made for chain end effects. However, we will follow the practice here of determining η_0 by extrapolating to higher molecular weights along a line of unit slope. We must remember, though, that this procedure may lead to a value of η_0 which is smaller than the proper value.

It is easily seen that the sums of Eq. (4) converge so rapidly that one need not carry more than the first few terms of each series. For this reason, the upper limits of the series, N and σ, are not needed in practice. Therefore, we observe that all of the quantities of Eq. (4) are measurable and so the relation may be used to predict viscoelastic behavior provided one knows M_c/M, η, η_0, and ν. We shall test the theoretical result against experiment in a later section.

In deriving Eq. (4) we have assumed a uniform molecular-weight polymer. The effect of molecular weight distribution is discussed in Appendix II. It should also be mentioned that Ferry *et al.*[1] have given a treatment of the effect of entanglements which is similar in many respects to the one we have given here. Some of the qualitative aspects of the effects of entanglements have been previously pointed out by A. V. Tobolsky among others.

To illustrate the various terms of Eq. (2) we have taken the values which apply for a PMMA of $M=652\,000$ at $T=140$°C. From the data of Table I we find that $\eta=10^{7.72}(65.2/6.30)^{3.4}=1.5\times10^{11}$ poise. Similarly, we find that the viscosity of a polymer having $M=M_c=10\,500$ is 1.32×10^5 and so $\eta_0=1.32\times10^5(65.2/1.05)=8.0\times10^6$ poise. These facts allow us to calculate the three terms of Eq. (4) and these are shown in Fig. 4.

The curve labeled I is the network response and appears to occur almost instantaneously at this temperature. Curve II is the beginning of the elastic response of the chain as a whole. Actually the curve continues upward until it reaches a value $(2M/3M_c)=40$ times higher than curve I at which point it, too, becomes flat. Curve III on the other hand represents the true viscous flow and it will continue upward in an exponential fashion. The solid line is the sum of these three types of extension and represents the observed extension of the sample.

If at some time after curve II has reached its limit, the load is removed, one will then observe an elastic recovery of the sample. The recovery curve will be the sum of curves I and II where these curves are now supposed to represent contractions of the sample.

In a purely qualitative sense, the sample at the temperature for which Fig. 4 is drawn will appear to be a rather rubbery material although it would be considered fairly "dead." By increasing the temperature, the whole curve is shifted towards the left and what previously occurred in 100 minutes may now occur in one minute. For example, if the temperature is raised so that point B is at $t=1$ min, then the material will appear very elastic. If point B is shifted to about $t=0.01$ min, the sample will appear to deform irreversibly and will then be in the molding range. On the other hand, if the temperature is lowered to about 120°C, point A will be reached in 1 minute and the polymer will appear quite hard. At 110°C it will actually be brittle.

In order to show the effect of molecular weight we have computed similar curves for M equal to 2.04, 6.52, and 41×10^5. These are shown as the solid lines in Fig. 5. It will be observed that for a molecular weight of 204 000 the elastic effects have become nearly overshadowed by true viscous flow even though there are still about 40 entanglement points per chain.

In order to obtain the bulk viscosity of polymers from a tensile creep and recovery curve one need only carry out a measurement at a temperature high enough so that a fair amount of measurable flow occurs during the time of the experiment. The original data may then appear somewhat like that shown in Fig. 6. Actually, measurements are carried out for times about 10 times longer than shown. This is necessary since in this type of experiment it is impractical to obtain the maximum elastic elongation of the chain and so the

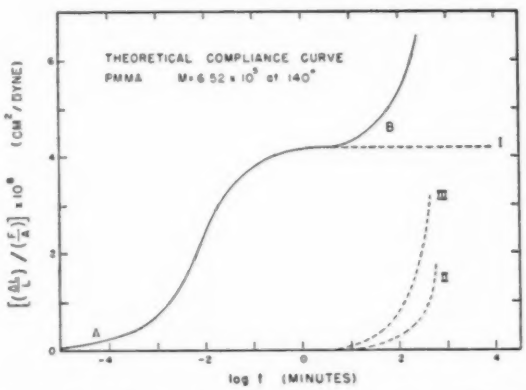


FIG. 4. Compliance of a PMMA having $M=6.52\times10^5$ at 140°C as calculated from Eq. (4). Curves I, II, and III represent the response of the entanglement network, the chain as a whole and the true-viscous flow respectively.

elastic recovery curve begins to deviate from the elastic creep curve at times approaching the maximum loading time.

One may calculate the polymer viscosity by two methods from the data acquired. The most reliable method is to subtract the recovery curve from the creep curve. The straight line so obtained represents the true viscous flow under the applied load. One then has that the tensile viscosity is given by

$$\eta = (F/A)/[d(\Delta L/L)/dt] \quad (5)$$

where F/A is the tensile force per unit cross-sectional area.

A second method is to allow the sample to recover as completely as it will. For an experiment where the nonrecoverable elongation was L when the load was applied for a time t, one has

$$\eta = (F/A)/[(\Delta L/L)/t]. \quad (6)$$

Good agreement was found between these two methods.

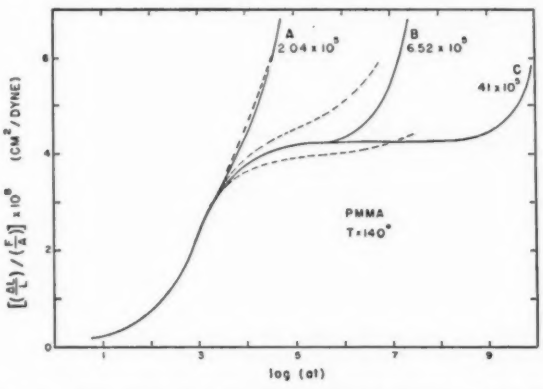


FIG. 5. Theoretical (full lines) and experimental (dashed lines) compliance for PMMA fractions having molecular weights of 2.04, 6.52, and 41×10^5.

EXPERIMENTAL TEST OF THEORY

In practice it is not feasible to carry out a creep measurement over a time interval of 10^{10} minutes and so one is unable to observe the full creep curves shown in Fig. 5. However, we see from Eq. (4) that the time, t, always enters in the form t/τ_c where

$$\tau_c = (4\eta_0/\pi^2\nu kT)(M_c/M)^2.$$

We therefore observe that τ_c is a characteristic elastic response time for the plastic and it is approximately the time needed for the polymer to complete $\frac{2}{3}$ of its network type response. It is therefore of value to plot the polymer response against $\log(t/\tau_c)$ rather than against $\log(t)$ because on the former plot the curve for a particular polymer will be nearly independent of temperature. A small variation will be introduced by the factor $(1/\nu T)$ in the first two terms of Eq. (4) but since this effect is hardly larger than experimental error, we shall neglect it for the present purposes.

It therefore follows from our theoretical concepts

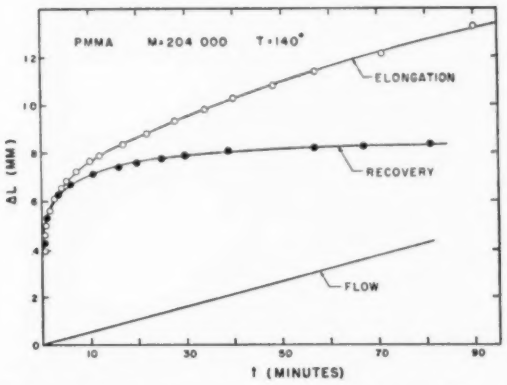


FIG. 6. A typical response curve for a PMMA having $M = 204\,000$.

of the process involved and as expressed in Eq. (4), that a change in the temperature at which the experiment is carried out causes a change in τ_c and is therefore equivalent to a shift in experimental time scale. Although, as we have indicated previously, it is logical to plot the experimental data *vs* $\log(t/\tau_c)$, the common practice is to define an arbitrary parameter, $a\sim1/\tau_c$, and plot all data against $\log(at)$. This practice was introduced by Ferry[10] and Tobolsky[11] who were among the first to recognize the interrelation between temperature and time scale. We shall bow to common usage and follow that notation here also. However, we should keep in mind that $(1/a)$ has an important physical significance and is proportional to the characteristic response time as defined above. Typical experimental data are shown in Fig. 7 where we have plotted the observed creep curves for a fraction of PMMA having $M = 652\,000$. Actually, curves were also obtained at other temperatures but they have not been included in this figure. The curves shown in Fig. 7 for various temperatures may be made to overlap by shifting the curves along the horizontal axis. Such a shift is equivalent to multiplying t by a constant and is just the type

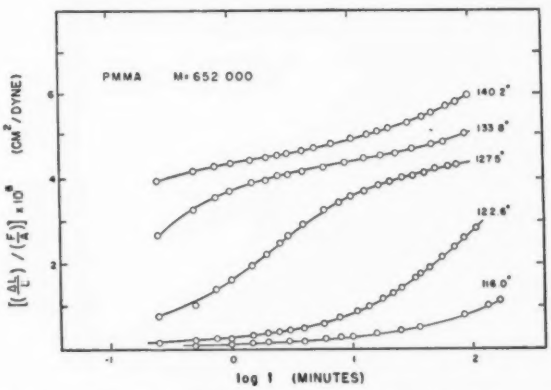


FIG. 7. Typical tensile creep data for a PMMA having $M = 652\,000$. The temperatures at which the data were taken are indicated on the curves.

[10] J. D. Ferry, J. Am. Chem. Soc. **72**, 3746 (1950).
[11] See, for example, R. Andrews and A. Tobolsky, J. Polymer Sci. **7**, 221 (1951).

FIG. 8. Experimental compliance curves for PMMA fractions of $M=2.04$, 3.65, 6.5, and 41×10^5. The scale is chosen such that $a=1$ min^{-1} at 110.6°C.

of procedure which was shown to be justified by Eq. (4). After carrying out this temperature-time superposition, one arrives at the experimental curves of Fig. 8.

In Fig. 8 we have plotted the experimental creep curves for four different PMMA fractions. The horizontal coordinate is so chosen that if $a=1$ min^{-1} then the curve is the one actually observed at 110.6°C. (No experimental points are shown on these curves since they are much too numerous to include on such a small scale. The original data are all about as accurate as those shown in Fig. 7.)

In order to ascertain the validity of the theoretical formulation of the creep curve as expressed by Eq. (4), we have compared theory and experiment in Fig. 5. The broken lines represent the experimental curves obtained for PMMA fractions with molecular weights of 2.04×10^5, 6.52×10^5, and 41×10^5 while the solid curves were calculated for these molecular weights using Eq. (4) described previously. Since a very small change in temperature results in a large horizontal shift of these curves, it is difficult to obtain an exact value of the viscosity at the temperature for which the curves are plotted. However, within the experimental error, the lower portions of the theoretical and experimental curves were found to coincide.

The small but significant divergence between theory and experiment in the plateau region is probably the result of three factors. First, the method we have used to find η_0 is not very precise and if η_0 is too low, as might be expected from the aforementioned discussion of the extrapolation procedure, then the onset of the viscous flow portions of the theoretical curves will not occur soon enough. This appears to be the case. However, the experimental curves have not been extended far enough to be sure of this. Second, in the theoretical calculation, no provision was made for the fact that there must actually be a distribution of chain entanglement lengths. This, together with the third factor, the neglect of various small terms in the differential equation, will tend to make the theoretical curve appear too flat in the plateau region. On the whole we believe the agreement actually observed provides a fairly good confirmation of our theoretical picture.

TEMPERATURE DEPENDENCE

Very often we are interested in flow phenomena at temperatures other than 110.6°C. Since a change of temperature can be accounted for by changing the value of the time scale constant, a, we can best describe the temperature variation of the creep curve by tabulating values of a at different temperatures. This is done in Table II and Fig. 9. These values were obtained in the following way: if at a temperature T the creep curve developed 10 times faster than at 110.6°C, then $a=10$ at this temperature. As an example of the use of these values, suppose it is desired to find how a sample will behave at 140°C. From the graph we notice that $\log a_{140}=7.6$ and so we conclude that we must subtract 7.6 from the $\log t$ scale of Fig. 8 if the curve is to apply at 140°C.

Another feature of common interest is the apparent activation energy for viscous flow. From what has been previously said, it follows that this is essentially the same as the apparent activation energy for elastic response. Denoting this quantity by E as was done in the discussion of viscosity, it follows that since $\eta=\eta_{110}/a$ and since

$$E=-R[d(\ln\eta)/d(1/T)]$$

we have

$$E=R[d(\ln a)/d(1/T)].$$

This quantity is plotted in Fig. 10.

A similar variation of E with temperature has been reported previously by McLoughlin and Tobolsky for PMMA.[12] These authors used stress relaxation methods to obtain their data. Although they too observed a maximum in the curve at almost the same temperature where we have found one, their curve appears to peak more sharply than ours. Since their data at high temperatures are not very complete, the difference observed in that region is not too surprising. However, the difference at lower temperatures is probably real and we believe it to be a result of the particular experimental procedure followed in conditioning the test samples. In order to see how this can be true, we must first say a few words

TABLE II. Variation of the time scale factor with temperature for pure polymethyl methacrylate.

Temperature °C	60	70	80	90	100	110	120	130	140	150
$\log_{10}a$(min^{-1})	−9.10	−8.22	−7.00	−5.35	−3.20	−0.13	3.55	5.85	7.55	8.68

[12] J. McLoughlin and A. Tobolsky, J. Polymer Sci. 8, 543 (1952).

about the molecular mechanism responsible for the observed behavior.

A tentative mathematical explanation for the variation of apparent activation energy for viscous flow near the glass temperature has been given elsewhere.[13] It will be sufficient for the present purposes to review a qualitative picture of the phenomenon first presented by T. G. Fox and P. J. Flory[14] upon which the treatment of reference 13 was based.

It has been shown that the major temperature dependence of the viscosity of amorphous high polymers is a result of the fact that the viscosity is inversely proportional to the jumping frequency of the individual polymer segments. At high temperatures the local liquid structure of the polymer is very loose and so only a small energy, E_0, is needed in order for a segment to move. As T is lowered, the liquid structure tightens and E_0 becomes larger. However, when the glass temperature is reached, the time scale of the molecular rearrangement process becomes comparable to that of the experiment and so further tightening of the local liquid structure is dependent upon the rate at which the polymer is cooled.

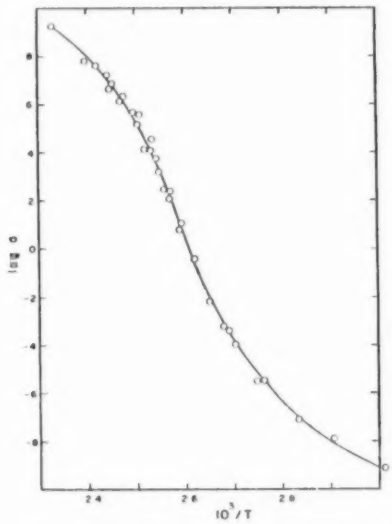


FIG. 9. Variation of the parameter a with temperature for PMMA.

To put this on a more quantitative basis, assume a fictitious model in which a polymer segment can only oscillate between two positions. We will also assume that there is just one mechanism by which the segment can move from one position to the other and that an energy E_0 is needed to execute this motion. The relative number of segments which possess enough energy to move will be

$$N/N_0=(1/kT)\exp(-E_0/kT).$$

This quantity is obviously proportional to the jumping frequency of any single segment.

[13] F. Bueche, J. Chem. Phys. **21**, 1850 (1953).

[14] See reference 4. I am indebted to T. G. Fox for discussions clarifying his published remarks on this question.

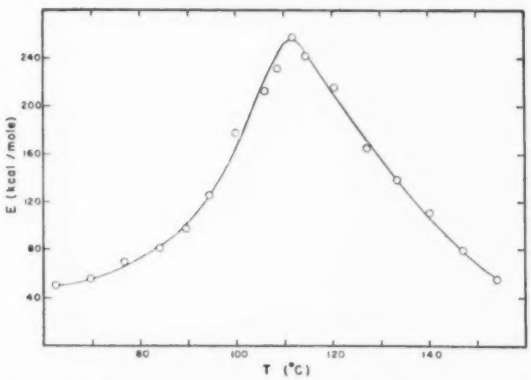


FIG. 10. Variation of the apparent energy of activation with temperature for PMMA. This polymer has $T_g=105°C$ and $T_c=135°C$.

If for simplicity we ignore the slow-temperature variation of the coefficient of the exponential we have for the apparent energy of activation for viscous flow

$$E=-k[d(\ln\eta)/d(1/T)]$$
$$=E_0+(1/T)[d(E_0)/d(1/T)].$$

From this it is obvious that near the glass temperature, where E_0 is increasing rapidly because of the tightening of the local liquid structure, one will observe a very large apparent energy of activation. At very high or very low temperatures, E_0 will be changing much less rapidly with T and so E will be more nearly equal to the actual energy involved in the jumping process.

The exact values of E at low temperatures will depend upon the amount of annealing of the sample. Since in this work we cooled our polymer through the glass temperature and down to 60°C at a fairly uniform rate over a period of about three weeks, it is apparent that we should expect a less rapid decrease in E than did McLoughlin and Tobolsky whose cooling cycle was much faster. We actually observed that even with the same sample the value of a could be shifted widely depending upon how quickly the sample was cooled.

We might also remark that a second reason for the decrease in E below T_g has been given by W. Kauzmann and others. It involves the assumption that the mechanism of molecular motion actually changes with temperature above and below T_g. Basically such a concept has much in common with the ideas presented here for the behavior observed above T_g and it is essentially just another way of describing the change in local liquid structure. On the other hand, below T_g the two explanations differ basically. It may well be that both reasons for the decrease in E below T_g are operative. As yet, no valid method of deciding between these two alternatives has been presented as far as we know.

The position of the maximum in the activation energy curve is also of interest. It is known from careful volume expansion measurements carried out in this

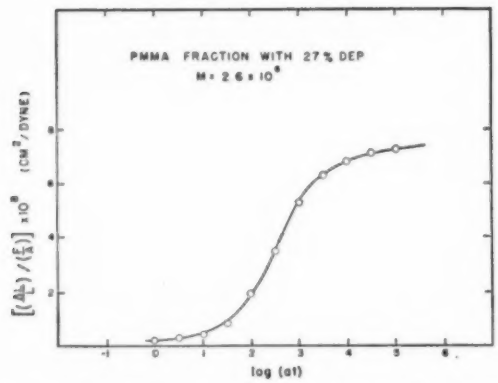


FIG. 11. Experimental compliance for a PMMA fraction ($M=2.6\times10^6$) plasticized with 27 percent DEP (points). For this system $T_g=19°C$, $T_c=55°C$, and $a=1$ min^{-1} at 39°C. The solid line represents the reduced curve for pure PMMA with $M=4.1\times10^6$.

laboratory by S. Loshaek that the glass temperature of pure PMMA is 105°C.[15] This is about 5° lower than the temperature at the peak of the activation energy curve. We must therefore conclude that the annealing process is already somewhat incomplete even at temperatures 5° above T_g as determined dilatometrically. This is probably not too surprising since it has long been known that a certain amount of rounding off occurs in the volume expansion data near the temperature taken for T_g.

EFFECT OF PLASTICIZER AND SIDE CHAINS

Our picture of the viscoelastic properties of the poly methacrylates cannot be complete until we understand the effects of plasticizer, copolymerization, and different side groups upon the tensile creep curve. In an effort to clarify these factors we have obtained tensile creep data for PMMA plasticized with DEP and with tricresyl phosphate. The tensile creep curves for polybutyl methacrylate (PBMA) and for a copolymer

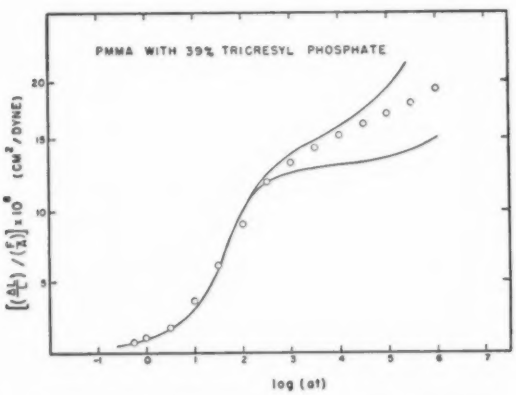


FIG. 12. Experimental compliance for a high molecular-weight PMMA containing 39 percent tricresyl phosphate (points). For this system, $T_g=-7°C$, $T_c=36°C$ and $a=1$ min^{-1} at 25°C. The solid lines represent the reduced curves for pure PMMA with $M=41$ and 6.5×10^5.

[15] S. Loshaek, J. Polymer Sci. **15**, 391 (1955).

containing 40 percent methyl methacrylate and 60 percent decyl methacrylate were also measured. Actually the decyl methacrylate was a mixture of about 50 percent octyl and 50 percent decyl methacrylate with small amounts of the other high methacrylates present. The polymer formed from this monomer alone had $T_g=-23°C$ as determined dilatometrically. The data for these polymers are shown in Figs. 11, 12, 13, and 14.

These widely different sets of measurement conditions can be unified if we notice they all have one thing in common, i.e., the backbone chain of the polymer is essentially the same in all cases. It is true that the butyl and decyl methacrylates have different side groups attached to the chain than does the methyl ester. However, since the groups concerned are far from being rigidly held to the chain, we might expect that their behavior will not seriously alter the general motion of the chain backbone and it is this motion which determines the major bulk properties of the

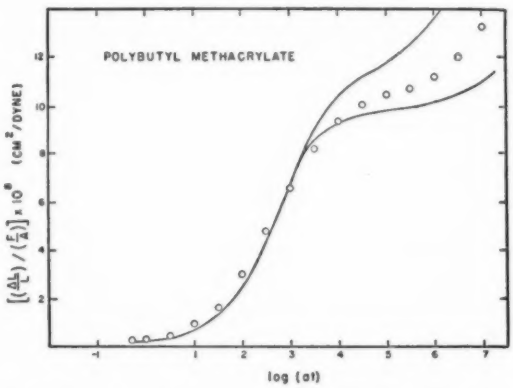


FIG. 13. Experimental compliance for a high molecular-weight polybutyl methacrylate (points). For this polymer $T_g=8°C$, $T_c=46°C$, and $a=1$ min^{-1} at 31°C. The solid lines represent the reduced curves for pure PMMA with $M=41$ and 6.5×10^5.

polymer. We shall in fact find that the side groups apparently act much as though they were plasticizer molecules and not attached to the chain at all.

From Eq. 4. it may be seen that the shape of the tensile creep curve (on a log$t$ plot) is dependent only upon the coefficients multiplying the three terms. In particular, the coefficient of the first term, representing the response of the network formed by interchain entanglements, is equal to $(4M_c/3\pi^2\nu kt)$. If one plasticizes the polymer, the average molecular weight between entanglements, $M_c/2$, will increase as argued previously. That is, if one has a mixture which contains 50 percent by volume of plasticizer, one would expect M_c to have twice the value it has for the pure polymer since there will now only be half as many molecules per unit volume to become entangled.

We shall assume tentatively that additional chain side groups may be considered to influence M_c in the same way as plasticizer. For example, methyl

methacrylate has a molecular weight which is (1/2.1) times smaller than that of our "decyl" methacrylate. Therefore, if we take the density of polydecyl methacrylate and PMMA to be the same, then M_e for polydecyl methacrylate should be 2.1 times the value found for PMMA. The quantity $M\nu$ is equal to $6\times10^{23}\times\rho$ where ρ is the concentration of polymer in the mixture in gm/cc.

This means, if our assumption concerning the effects of side groups and plasticizer is correct, the shapes (but not the positions on the time scale) of the curves of Figs. 11, 12, 13, and 14 should be the same as for PMMA except that the vertical scale of the PMMA curve must be multiplied by $(\rho_0T_0/\rho T)\times(M_e/M_{e0})$ in each case. The quantities without subscript refer to the polymer under consideration whereas the subscript, zero, indicates the value for pure PMMA. Since the viscosity of the mixtures will be different in each case, the log$t$ scale will be shifted for each curve indicating a different value for a in each case. Also, the length of

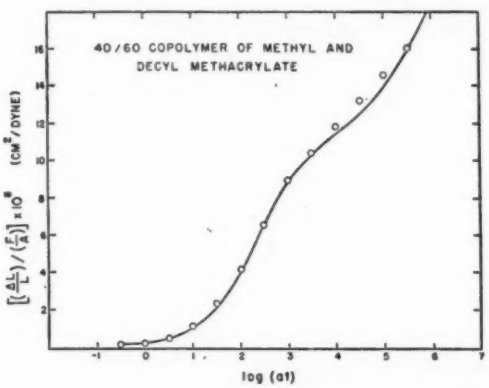


FIG. 14. Experimental compliance for a 40/60 copolymer of methyl and "decyl" methacrylate (points). For this copolymer $T_g=10°C$, $T_c=52°C$, and $a=1$ min^{-1} at 37.5°C. The solid line represents the reduced curve for pure PMMA with $M=3.65\times10^5$.

the plateau region will vary with variations of M_e/M and so the molecular weight of the polymer will appear to decrease with increasing plasticizer content.

We have checked these ideas by comparing the observed experimental curves with the curves for the pure PMMA which have been multiplied by the factor $(\rho_0T_0M_e/\rho TM_{e0})$. The value of M_e in each case was calculated assuming that added side groups and plasticizer increase the molecular weight between entanglements as outlined above. The modified curves for PMMA are indicated by the solid curves in Figs. 11, 12, 13, and 14. In every case we find excellent agreement between the expected and experimental curves. It therefore appears that our basic assumption, i.e., that distant pendant groups and added plasticizer serve only to dilute the polymer and do not otherwise affect the shape of the response curve, is valid in the present region of measurements.

However, added side groups and plasticizer seriously

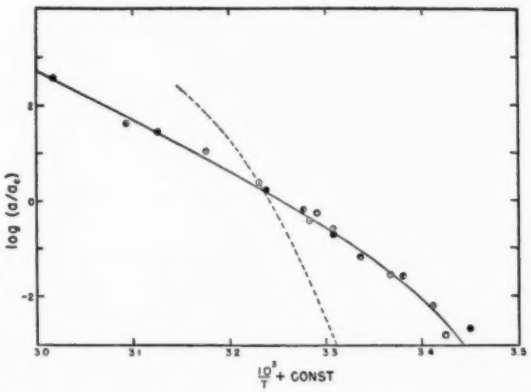


FIG. 15. Variation of the parameter a with temperature for various systems: ◐ PMMA plasticized with TCP, ● polybutyl methacrylate, ⊙ copolymer of methyl and "decyl" methacrylate, ○ PMMA plasticized with DEP. The dotted line is for pure PMMA.

affect the temperature behavior of the polymer. As an arbitrary method of describing this effect, we shall denote by T_c that temperature at which the polymer will have a value of $(\Delta L/L)/(F/A)=M_e/12\nu MkT$ at a time of one minute after the load is applied. This value of the compliance was chosen since it is 0.5 of the plateau value. These values are given in the legends for Figs. 11, 12, 13, and 14.

As we did in the previous graphs for pure PMMA we can define a parameter, a, such that when a master creep curve is plotted against log(at), the experimental data may be fitted to this by a proper choice of the parameter. We will take $a=a_c$ min^{-1} at $T=T_c$. It is then possible to make a plot of log(a/a_c) *vs* $1/T$ for each one of the materials of Figs. 11–14. Since the value of T_c is about the same for all four of these materials (ranging from 35 to 55°C), we might expect a to vary in much the same way for all of the materials. The data are shown in Fig. 15 and it is seen that, within experimental error, the above assumption is correct.

Activation energies for the viscous and elastic response of these materials can be obtained from the slope of this curve as shown previously. For comparison, we have included the similar plot for the pure polymer as the broken line of Fig. 15. It is seen that the activation energies for these materials which soften at a temperature some 100° below the corresponding temperature for pure PMMA are about one-third as large as for PMMA.

This decrease in activation energy with increased plasticization is to be expected from the considerations previously given.[13] It has been shown that the activation energy will be large for a polymer segment which requires a movement of a large number of other segments in order that it may move. As one adds plasticizer, it seems fairly obvious that the presence of small disconnected molecules (or fairly loosely held chain side groups) will allow a chain segment to move

without disturbing as large a region of the polymer as was previously necessary. This effect is almost certainly the cause of the decreased activation energy observed in this case.

ACKNOWLEDGMENTS

The author is deeply indebted to Dr. T. G. Fox for many helpful suggestions during the course of this work. The aid of H. Nields and J. Shetter in obtaining part of the data presented in this report is gratefully acknowledged. J. Kinsinger furnished three low molecular-weight fractions used in this work. We wish to thank him for these and also for his helpful comments.

APPENDIX I

In a previous paper[2] we have presented a method for calculating the response of a polymer in a tensile creep experiment. The method requires a knowledge of the forces acting upon the segments of the polymer chain. We assumed at that time that these forces were distributed smoothly along the chain and it turns out that this assumption does not invalidate the treatment at long times after application of the load, i.e., after viscous as well as elastic flow has become important. However, the initial deformation is sensitive to the presence of positions of localized stress along the chain resulting from chain entanglements. We will extend the previous treatment to include such effects in this appendix.

It has been shown that the average force on a chain segment is the force required in the calculation provided the force is independent of the velocity of the segment, ξ_i. However, since entanglement forces which retard the motion of the chain depend upon ξ_i, it is necessary to modify the calculation so as to include them.

Let us assume that each chain has N-segments with $2N/\sigma$ chains entangled with it. (We will neglect end effects in this discussion. This is allowable since N/σ will be large in cases where the effects of entanglements become noticeable.) Half of the entangled chains will be acting to accelerate the molecule under consideration. It turns out that these can be handled as was done previously. The remaining N/σ entanglements will result in making segments at entanglement points appear to have an abnormally large friction factor, f, in comparison to the normal friction factor, f_0. For convenience we shall take the retarding entanglements to be located at segments numbered (calling one end segment zero) $\sigma/2$, $3\sigma/2$, $5\sigma/2$, etc.

These ideas may be combined into the calculation previously presented. When this is done, the differential equation for the generalized coordinates becomes

$$(f_0/m)\dot{q}_n+(2f/mN)\sum_j\sum_{p=0}^{N}\dot{q}_p\cos(jk_p)\cos(jk_n)+\omega_n^2q_n=F_n, \quad \text{(A1)}$$

where

$$F_n=(FR_x/2N\eta)(2/mN)^{\frac{1}{2}}[f_0+(f/\sigma)][N+(4N^2/n^2\pi^2)] \quad \text{for } n \text{ odd}<N,$$
$$F_n=0 \quad \text{for } n \text{ even}<N,$$

and the other symbols are the same as in reference 2. The sum over j is to be taken over $j=\sigma/2, 3\sigma/2, \cdots, N$.

It will be noticed that we have used R_x, the x-component of the mean square end to end distance of the chain. In reference 2 we used R instead of R_x and as a result the answer given there is too large by a factor of three. (The network calculation of reference 2 is not affected by this.) I am indebted to Professor J. D. Ferry for pointing out the fact that my previous result was probably too large.

For convenience we define

$$a_{pn}=(2f/Nf_0)\sum_j\cos(jk_p)\cos(jk_n). \quad \text{(A2)}$$

It would be a formidable task to solve the above differential equation exactly. However, to a good approximation we can put

$$\begin{aligned} a_{pn}&=0 && n\neq p\\ &=0 && n=N/\sigma, 3N/\sigma, \cdots\\ &=f/\sigma f_0 && n=p\neq N/\sigma, 3N/\sigma, \cdots. \end{aligned} \quad \text{(A3)}$$

With these values for a_{pn}, the solution of Eq. (A1) becomes a simple matter and we can then obtain q_n. After that, the displacement of the ith chain segment, ξ_i, may be found from the relations given in reference 2. It is found that

$$\xi_i\sim\sum_n F_n\{1-\exp[-t/\tau_n]\}\cos(ik_n)+\sum_p F_p\{1-\exp[-t/\tau_p]\}\cos(ik_p) \quad \text{(A4)}$$

where

$$p=N/\sigma, 3N/\sigma, \cdots, N$$
$$n=1, 3, 5, \cdots, N$$

and where

$$\tau_n=(\eta/\eta_0)\tau_p=(4\eta/\pi^2\nu kTp^2).$$

We have denoted the observed viscosity of the polymer by η. If no entanglements existed, the viscosity would be smaller by a factor of $(1+f/f_0\sigma)^{-1}$ and this viscosity we denote by η_0.

The form of this relation shows clearly that the response is the sum of that resulting from the stretching of the entanglement network and that which we had previously calculated. We could calculate the proper correspondence to a bulk sample by the use of energy considerations, but since we already know the answer for the two component cases we can write down the end result at once. After adding on the viscous-flow

term and after redefining n and p we obtain

$$(\Delta L/L)/(F/A) = (4\sigma/3\pi^2\nu NkT)\sum_{p=1}^{\sigma/2}(2p-1)^{-2}[1-\exp(-t/\tau_p)] + (32/3\pi^4\nu kT)\sum_{n=1}^{N/2}(2n-1)^{-4} \times[1-\exp(-t/\tau_n)]+t/\eta, \quad \text{(A5)}$$

where p and n are integers and

$$\tau_p=4\eta_0\sigma^2/(2p-1)^2\pi^2\nu kTM^2,$$
$$\tau_n=4\eta/(2n-1)^2\pi^2\nu kT.$$

APPENDIX II

It is of interest to consider the effect of molecular weight distribution on the creep curves of plastics. We have already showed that Eq. (A5) [or Eq. (4)] applies for a sharp fraction in which all the molecules have approximately the same molecular weight. In the event that such is not the case, we see at once that the elastic response resulting from the stretching of the entanglement network (term 1 in Eq. (A5)) is only displaced in time scale and not otherwise altered provided the molecular weight remains high.

The variation of the viscosity term with molecular-weight distribution has not been worked out in detail but it has been shown experimentally by Fox and Flory that it varies with M_w as expected from Fig. 3.[4,5,7]

The dependence of the second term, the stretching of the polymer molecule as a whole, upon molecular weight distribution is rather complex. Ferry *et al.* have derived a relation which shows how the value of this term depends on molecular weight when $t\rightarrow\infty$. They approached the problem from the standpoint of relaxation time distributions.[16] We will show here how the effects of molecular-weight distributions may be calculated using our approach to viscoelastic phenomena. Although we shall restrict the treatment to the case of tensile creep, application of the method to other problems is not far different.

Since we have already discussed the behavior of the entanglement term in Eq. (4) we will only consider the second term which was treated in some detail in reference 2. If we consider the pth species (i.e., those molecules having a molecular weight M_p) the treatment is exactly the same as in reference 2 down through the derivation of the equation for the elongation of the molecule which is

$$\Delta x_p=(4\mathfrak{F}_p/N_p^2m)\sum_{n\ \text{odd}}(1/\omega_{np})^2 \times(2N_p^2/n^2\pi^2)[1-\exp(-t/\tau_{np})],$$

where

$$\mathfrak{F}_p=R_{xp}Ff/\eta. \quad \text{(A6)}$$

We shall assume a unit cube sample. However, we cannot say as was done for a fraction that $\Delta L/L=\Delta x/R_z$ in the present instance since there are many values of $(\Delta x/R_z)$ as a result of the molecular-weight distribution. Instead we use the fact that

Energy stored in chains = Energy stored in sample

$$W_c=W_s,$$
$$\tfrac{1}{2}F\Delta L=\beta\sum_p\nu_p\mathfrak{F}_pN_p\Delta x_p, \quad \text{(A7)}$$

where ν_p is the number of chains per unit volume having $M=M_p$ and β is a pure number which can be found directly or by comparison with the answer previously found for a fraction. It turns out to be $1/24\sqrt{3}$.

Although it is possible to find out exactly how each point of the stretching curve depends on molecular weight, we will restrict the present treatment to the case treated by Ferry *et al.*,[16] the compliance for $t\rightarrow\infty$. In that instance we have from Eqs. (A6) and (A7) that

$$\Delta x_p=\mathfrak{F}_pN_p^2l^2/36(3)^{\frac{1}{2}}kT$$

and

$$W_c=(\beta F^2l^2f^2/18kT\eta^2)\Sigma N_p^3R_p^2. \quad \text{(A8)}$$

But we know from reference 8 that

$$\eta=(f/12)\Sigma\nu_pN_pR_p$$

which leads to the result that

$$(\Delta L/L)/(F/A)=(1/9kT)[\Sigma N_p^4\nu_p/(\Sigma N_p^2\nu_p)^2]$$

which simplifies to the result that

$$(\Delta L/L)/(F/A)=(M_{z+1}M_z)/(9\rho RTM_w),$$

where M_{z+1}, M_z, and M_w are the $z+1$, z, and weight-average molecular weights, respectively.

This is essentially the same result as is found employing the method used by Ferry *et al.*[16] The important feature to notice is that a very small percent of high molecular-weight material is extremely influential in determining the ultimate elastic compliance.

[16] Ferry, Jordan, Evans, and Johnson, J. Polymer Sci. (to be published).

Experimental Examination of the Statistical Theory of Rubber Elasticity. Low Extension Studies*

F. P. BALDWIN,† J. E. IVORY,‡ AND R. L. ANTHONY
Physics Department, University of Notre Dame, Notre Dame, Indiana
(Received October 6, 1954)

An experimental study was made of the behavior of two rubbers, Butyl (GR-I-18) and Paracril-35, in the low extension region (0 to 15 percent), at temperatures ranging from −40°C to +60°C. The results have been compared with the Guth-James theory of rubber elasticity. In general, good agreement with theory was observed. Two effects predicted by the theory have been confirmed, namely, the nonlinear dependence of stress upon temperature at constant extended length, and the shift of the thermoelastic inversion extension to higher values with increasing temperature. Small deviations, which increased with increasing extension and with decreasing temperature, were observed.

INTRODUCTION

As part of a series of experimental investigations designed to provide a critical evaluation of the Guth-James theory of rubber elasticity,[1] a careful study of the behavior of two rubbers at low extensions was undertaken. The low extension region was chosen for this initial study because of the interesting behavior of rubbers in this region. As is well known, rubbers exhibit a thermoelastic inversion in the vicinity of 5 to 7 percent extension. This phenomenon was first observed by Joule[2] in 1859, while engaged in measuring temperature changes accompanying the adiabatic stretching of rubber. Joule noted a slight cooling at low extensions, followed by a rise in temperature at higher extensions. These observations, along with the thermodynamic analysis of Kelvin,[3] indicated that for rubber specimens held at constant extended lengths, the temperature coefficient of stress must be negative at low extensions and positive at higher extensions. The first direct observation of this change in sign of the temperature coefficient of stress was made by Meyer and Ferri[4] in 1935. This was confirmed by the experimental work of Anthony, Caston, and Guth,[5] and of Peterson, Anthony, and Guth[6] in 1942. Later experimental work by Wood and Roth,[7] Gee,[8] Treloar,[9] and others contributed much to the understanding of the physical behavior of rubbers. Wood and Roth investigated the effect of working at constant "extension ratio," rather than at constant length, upon the stress-temperature relations of rubbers. Gee's early work was concerned primarily with the attainment of better equilibrium conditions through the use of a swelling technique. Treloar investigated deviations from the theoretically predicted stress-strain curve at moderate to high extensions.

All of the stress-strain studies mentioned above lacked the precision necessary to test adequately the predictions of the Guth-James theory in the region of low extensions. Accordingly, the present study was undertaken for the specific purpose of testing the theory in this region. The rubbers selected for this study were Butyl (GR-I-18) and Paracril-35. Butyl was chosen because of its relative stability to oxidation in comparison with more highly unsaturated polymers. The GR-I-18 used in this work is an isobutylene-isoprene copolymer having a mole percent unsaturation of 1.38. For purposes of comparison and contrast, a second rubber having relatively strong intermolecular forces and a fairly high second-order transition temperature was desired. Paracril-35, a copolymer of butadiene and acrylonitrile containing about 35 percent of the latter, was selected for this purpose. Extensions were limited to the ranges 0 to 10 percent for the Butyl rubber, and 0 to 15 percent for the Paracril rubber. The temperature ranges covered were from −40°C to +40°C for the Butyl rubber, and from −20°C to +60°C for the Paracril rubber.

THEORY

The Guth-James theory[1] gives the following expression for the stress as a function of the extension, the temperature, and the pressure:

$$Z=KT\left[L-\frac{V(LTP)}{L^2}\right]. \qquad (1)$$

Here, $Z=F/A_0$ is the nominal stress, $L=l/l_0$ is the relative length, and $V=v/v_0$ is the relative volume.

* This work was supported in part by the Office of Naval Research. It represents portions of two theses submitted by F. P. Baldwin and J. E. Ivory to the Graduate School of the University of Notre Dame, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
† Holder of U. S. Rubber Company Fellowship during the course of part of this work. Now at Esso Laboratories, Linden, New Jersey.
‡ Now with Bell Aircraft Corporation, Buffalo, New York.
[1] H. M. James and E. Guth, J. Polymer Sci. **4**, 153 (1949).
[2] J. P. Joule, Phil. Trans. **149**, 91 (1859).
[3] W. Thomson (Lord Kelvin), Quart. J. Math. No. 1, April (1855); *Mathematical and Physical Papers* (Cambridge University Press, Cambridge, England, 1882), Vol. 1, pp. 309–10.
[4] K. H. Meyer and C. Ferri, Helv. Chim. Acta. **18**, 570 (1935).
[5] Anthony, Caston, and Guth, J. Phys. Chem. **46**, 826 (1942).
[6] Peterson, Anthony, and Guth, Ind. Eng. Chem. **34**, 1349 (1942).
[7] L. A. Wood and F. L. Roth, J. Appl. Phys. **15**, 781 (1944); and J. Appl. Phys. **15**, 749 (1944).
[8] G. Gee, Trans. Faraday Soc. **42**, 585 (1946).
[9] L. R. G. Treloar, Trans. Faraday Soc. **40**, 59 (1944).

The quantities A_0, l_0, and v_0 are, respectively, the cross-sectional area, the length, and the volume of the unstrained specimen at the reference temperature T_0. The quantities F, l, and v are, respectively, the tension force, the length, and the volume of the strained specimen at the test temperature, T. P is equal to the hydrostatic pressure of any fluid medium surrounding the specimen, and K is a constant assumed to be independent of both temperature and extension. The factor K, which according to the theory is related to the number of cross links per cubic centimeter of the rubber, plays the role of an adjustable constant in fitting the theoretical stress-strain curve to its experimental counterpart.

The definitions of Z, L, and V given above are equivalent to normalizing the initial sample dimensions to those of a 1 cm cube at $T=T_0$. Hence, we may also consider Z to be the total stretching force, and L the actual stretched length in centimeters, for such a model. This interpretation is convenient in the application of thermodynamics to the problem of rubber elasticity. When it is employed, the fundamental quantity of rubber to be considered in thermodynamic equations is then 1 cm³, and not 1 g.

For purposes of resolving the Guth-James force expression or the experimental stress-strain curves into their entropy and internal energy components, the following thermodynamic considerations are useful. Assuming that an equation of state exists for the rubber, and that satisfactory equilibrium conditions have been achieved, the combined first and second laws of thermodynamics state:

$$dU=TdS-PdV+ZdL \tag{2}$$

for the case of unilateral extension. Here U is the internal energy, and S is the entropy. Employing the Gibbs function $G=H-TS$ where $H=U+PV$ is the enthalpy, the following stress relation may be deduced:

$$Z=\left(\frac{\partial G}{\partial L}\right)_{TP}=-T\left(\frac{\partial S}{\partial L}\right)_{TP}+\left(\frac{\partial H}{\partial L}\right)_{TP}. \tag{3}$$

With the aid of an appropriate Maxwell relation, Eq. (3) may also be written in the form,

$$Z=T\left(\frac{\partial Z}{\partial T}\right)_{LP}+\left(\frac{\partial H}{\partial L}\right)_{TP}. \tag{3'}$$

It is convenient to refer to the first term on the right of Eqs. (3) and (3′) as the entropy component Z_S of the total stress Z and to the second term as the enthalpy component Z_H. Thus,

$$Z_S=-T\left(\frac{\partial S}{\partial L}\right)_{TP}=T\left(\frac{\partial Z}{\partial T}\right)_{LP} \tag{4}$$

and

$$Z_H=\left(\frac{\partial H}{\partial L}\right)_{TP}=Z-Z_S. \tag{5}$$

The application of Eqs. (4) and (5) to the theoretical model represented by Eq. (1) yields:

$$Z_S=KT\left[L-\frac{V(LTP)}{L^2}\right]-\frac{KT^2}{L^2}\left(\frac{\partial V}{\partial T}\right)_{LP} \tag{6}$$

and

$$Z_H=\frac{KT^2}{L^2}\left(\frac{\partial V}{\partial T}\right)_{LP}. \tag{7}$$

The factor $(\partial V/\partial T)_{LP}$ appearing in Eqs. (6) and (7) is equal to α_{LP}, where α_{LP} denotes the volume coefficient of thermal expansion of the rubber model, subject to the restrictions of constant length and pressure. For real rubbers the coefficient α_{LP} should be practically indistinguishable from the ordinary coefficient α measured without the restriction of constant length. Evidence that α is nearly independent of extension, for extensions ranging up to 200 percent, is supplied by the data of Copeland.[10] Accordingly, the same behavior is ascribed to the model, and α_{LP} is replaced by α. Further, Poisson's ratio for real rubbers is almost exactly $\frac{1}{2}$, the dependence of V upon L and P being practically negligible in comparison with its dependence upon T. Hence, $V(LTP)$ is replaced by $[1+\alpha(T-T_0)]$ in Eq. (6). Equations (6) and (7) thus become:

$$\begin{aligned} Z_S&=KT\left[L-\frac{1+\alpha(T-T_0)}{L^2}\right]-\frac{KT^2\alpha}{L^2}\\ &=KT\left[L-\frac{1+\alpha(2T-T_0)}{L^2}\right] \end{aligned} \tag{8}$$

and

$$Z_H=\frac{KT^2\alpha}{L^2}. \tag{9}$$

To within the same approximation the original stress, Eq. (1), may now be written:

$$Z=Z_S+Z_H=KT\left[L-\frac{1+\alpha(T-T_0)}{L^2}\right]. \tag{10}$$

Equations (8)–(10) are approximate expressions which correspond to the Eqs. (6), (7), and (1), respectively. They are also exactly the equations which one obtains for the incompressible analog of the original model, i.e., when Poisson's ratio is taken to be exactly one-half. We shall employ Eqs. (8), (9), and (10) for comparison with experimental results. They are of additional interest in that they predict two effects which had not been observed in earlier, less precise, experimental work.[5,6] First, Eq. (10) predicts a nonlinear dependence of Z upon T, at constant specimen length. Second, Eq. (8) predicts that the thermoelastic inversion extension (i.e., the extension for which $Z_S=0$) should depend upon temperature, and in fact,

[10] L. E. Copeland, J. Appl. Phys. **19**, 445 (1948).

should increase with increasing temperature, *viz*:

$$L_{\text{inv}}=[1+\alpha(2T-T_0)]^{\frac{1}{3}}. \tag{11}$$

One final point should be noted before proceeding to the experimental part of the work. For a real rubber at atmospheric pressure the enthalpy contribution to the stress, as given by Eq. (5), is indistinguishable from an internal energy contribution (see, for example Gee[8]). Thus for all intents and purposes we have $Z_H \simeq Z_U$, or

$$\left(\frac{\partial H}{\partial L}\right)_{TP} \simeq \left(\frac{\partial U}{\partial L}\right)_{TP}. \tag{12}$$

EXPERIMENTAL METHOD

The stress relaxation method was selected for obtaining the required equilibrium stress-strain curves. In this method the sample is first extended to a relative length L at the reference temperature T_0, this temperature being the highest temperature to be used in the experiment. The sample is then held at constant length and at temperature T_0 until the time rate of decay of stress becomes negligibly small, i.e., so small that it can have no appreciable influence on the following stress-temperature run. In the work reported here the time allotted for this initial relaxation process was usually about 12 to 15 hours. After this stage the temperature is lowered in steps, the length of the specimen still being held constant, and the stress Z is determined at each new temperature after thermal equilibrium has been attained. This procedure is continued until the lowest temperature desired has been reached. Although it is not essential, the temperature may then be raised in steps, and the identical procedure employed for increasing temperatures. Comparison of these data with those obtained for decreasing temperatures affords an excellent check on the degree of reversibility achieved, i.e., on the extent to which undesired time effects have been suppressed. This complete procedure was followed in obtaining all of the results reported here.

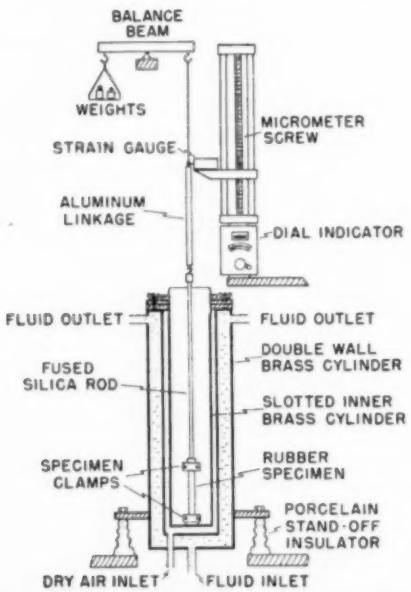


FIG. 1. Schematic diagram showing main features of apparatus.

By repeating the procedure just described for a series of samples, each sample being stretched to a different extended length, a family of equilibrium stress-temperature curves is obtained. Cross plotting these for various constant values of the temperature yields the equilibrium stress-strain curves.

The apparatus employed for obtaining the stress-temperature data is shown schematically in Fig. 1. The desired temperature within the cylindrical air bath was secured by adjustment of the temperature and rate of flow of the liquid (ethyl alcohol) circulating between the double brass walls of the thermal inclosure. Temperatures were measured with a copper constantan thermocouple placed near the center of the sample, and were recorded continuously. Dry air (H_2SO_4 dried) was brought to the required temperature and introduced into the test chamber through the inlet at the base. This provided good mixing in the test chamber, and also prevented the formation of frost around the fused silica rod during low temperature operation. When the system was properly adjusted, temperatures within the air bath could be held constant to within 0.1°C over the complete 80°C temperature range. It was found necessary to include the aluminum strip in order to compensate for changes in length of the supporting iron framework (not shown), which resulted from small changes in room temperature. A Statham strain gauge was used as a null detector, the gauge output being fed to a Brown Electronik recorder. The gauge could be driven up or down by means of the micrometer screw whenever the sample length was to be changed for a new run. During a given run the gauge was held at a fixed position. When necessary, changes in force detected by it were compensated by changing the weights W in the scale pan suspended from one side of the equal arm beam balance. With a gauge input of 2.4 volts, obtained from a storage battery, a gauge deflection of 1 mil resulted in an output change of 0.32 millivolts. For the 0 to 3 millivolt Brown Recorder employed, this corresponded to a deflection of 32 chart divisions. For the same gauge input the load sensitivity was about 5.8 chart divisions/g. Brown chart deflections as small as one-half division were easily detectable.

In order to maintain sample lengths constant and independent of temperature, it was necessary that the clamp spacing should remain constant for constant strain gauge output. Conversely, changes in the strain gauge output should accurately reflect changes in the clamp spacing and not be the result of expansions or contractions of some other part of the apparatus. To secure this condition, the inner brass cylinder is threaded at its top end and passes through a threaded cap which

can be firmly fixed to the top of the chamber by bolts and wing nuts. This cylinder is slotted to provide for good air circulation, and has the lower specimen clamp attached to its base. By trial and error the immersion depth of the cylinder was varied until its thermal changes almost exactly compensated those due to the double-walled cylinder, the procelain insulators, the clamps, etc. This was done by comparing the thermal expansion of a 4.31-inch strip of copper, mounted between the clamps, with the value to be expected from data reported in the literature.[11] As finally adjusted, agreement to within 3 percent was obtained over a 90°C temperature range. This represented a discrepancy of only 2×10^{-4} inches from the predicted thermal expansion, and was well within the limits of precision set by the dial and screw measuring device. Thus, for rubber samples of lengths 4 inches or greater, extended lengths could be held constant and independent of temperature to better than 0.01 percent. A second and indirect test of the final adjustment was the use of the apparatus to measure the thermal expansion coefficient of Butyl rubber at zero stress. The linear expansion coefficient obtained was 1.894×10^{-4}/°C, giving a volume coefficient of 5.68×10^{-4}/°C. This value compares very favorably with that reported by Bekkedahl,[12] namely 5.67×10^{-4}/°C for unvulcanized Butyl, measured with a volume dilatometer. Similar measurements for the Paracril-35 rubber gave a value of 6.65×10^{-4}/°C for its volume coefficient of thermal expansion.

The rubber specimens used in this study were died out of standard 6-inch square tensile sheets. The specimen shape was that of a square-headed dumbbell, previous tests having shown that this type of sample gave excellent correlation between percentage extensions computed from clamp separations and those computed from separation of dots placed on the central part of the specimen. The average dimensions of the neck of the specimen when unstrained were 4 in. by 0.25 in. by 0.075 in.

The rubbers had the following compositions and cures (amounts given in parts by weight):

Butyl rubber		Paracril rubber	
GR-I-18	100.00	Paracril-35	100.00
Zinc oxide	5.00	Zinc oxide	5.00
Sulfur	0.80	Stearic acid	1.00
Tuads	1.00	Sulfur	1.75
		Altax	2.00
Cured 75 minutes at 300°F		Cured 80 minutes at 287°F	

RESULTS

Figure 2 shows the dependence of stress upon temperature observed for the Butyl rubber at three different values of the percentage extension. [The percentage extension, as used in this paper, is defined as 100(L–1).] Both effects predicted by the theory

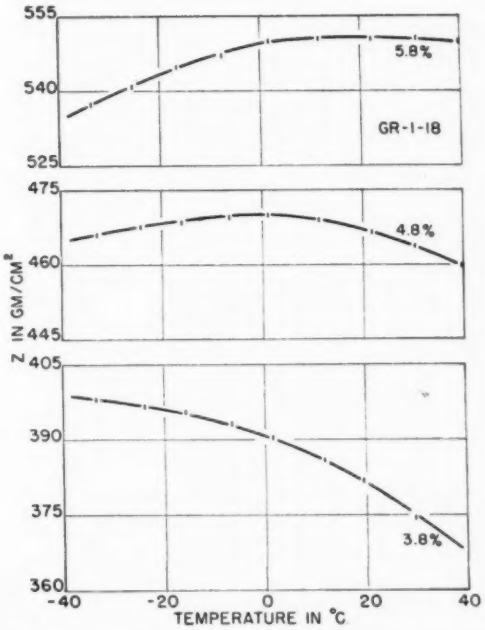


FIG. 2. Stress-temperature behavior of GR–I–18 for three selected values of percentage extension.

are confirmed by the curves shown here; the dependence of stress upon temperature is decidedly nonlinear, and the thermoelastic inversion temperature is shifted to higher values at higher extensions. Note the shift of the peak to the right as the extension is increased! This is the first time either of these two effects have been observed with certainty in this laboratory. Appreciable deviations from linearity were not apparent in earlier stress-temperature data,[5,6] owing both to a lack of sufficient precision in the measurements, and probably also to our failure to maintain extended lengths sufficiently constant during the stress-temperature runs. Curves similar to those shown in Fig. 2 were observed also for the Paracril rubber. The complete set of stress-temperature curves is shown in Fig. 3 for the Butyl rubber, and in Fig. 4 for the Paracril rubber. In order to cover the full range of extensions for each rubber in a single figure, the stress scales used in Figs. 3 and 4 have been considerably reduced over that employed in Fig. 2, with the result that deviations from linearity are now scarcely perceptible.

To evaluate the constant K appearing in the Guth-James theory, the "true stress" was plotted as a function of extension, at the reference temperature T_0, for each rubber. The required values of Z and L were obtained from the curves of Figs. 3 and 4 by cross-plotting. The true stress is defined as the tension force divided by the actual cross-sectional area (not the unstrained cross-sectional area). Making the assumption that Poisson's ratio is exactly 0.5 here, the true stress is then simply equal to the product of the nominal stress Z by the relative length L. This method of plotting yields a linear stress-strain curve at low

[11] F. C. Nix and D. MacNair, Phys. Rev. **60**, 597 (1941).
[12] N. Bekkedahl, J. Research Natl. Bur. Standards **43**, 145 (1949).

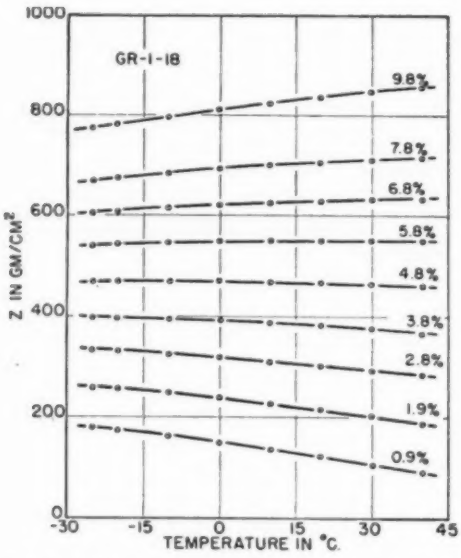


FIG. 3. Family of stress-temperature curves for GR–I–18.

extensions, as shown in Fig. 5. This enables a good determination of Young's modulus, E_0, to be made in the limit of $L=1$ and $T=T_0$. The values of E_0 obtained from the curves in Fig. 5 were 9.96×10^8g/cm² at 40°C for the Butyl rubber, and 20.4×10^8g/cm² at 60°C for the Paracril rubber. These values of E_0 were then equated to the theoretical expression $3KT_0$, obtained from Eq. 10 by differentiating with respect to L and putting $L=1$ and $T=T_0$. The values of K so obtained were 10.6 and 20.4g/cm²×°K for the Butyl and Paracril rubbers, respectively. All fitting of theory to experiment in this work was confined to the evaluation of K, as just described.

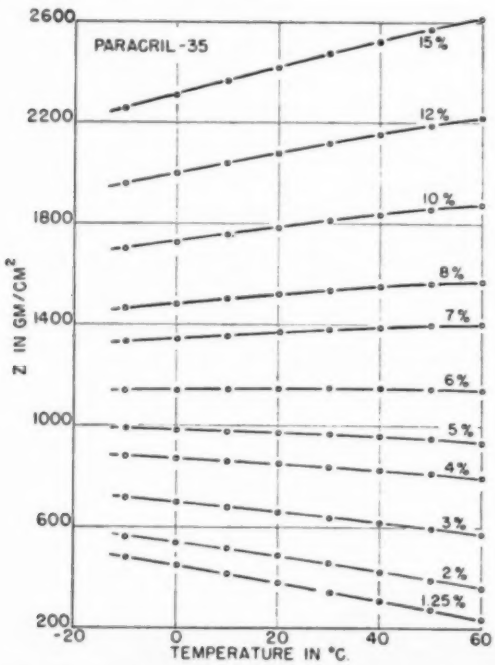


FIG. 4. Family of stress-temperature curves for Paracril-35.

The equilibrium stress-strain curves were obtained from the stress-temperature curves of Figs. 3 and 4 by cross-plotting at various fixed values of the temperature. The experimental entropy components were obtained by evaluating the point slope, $(\partial Z/\partial T)_{LP}$, as a function of temperature for each of the stress-temperature curves. The entropy component of the stress $Z_S=T(\partial Z/\partial T)_{LP}$ was then plotted as a function of temperature in order to achieve some smoothing. The resulting curves were cross-plotted at various fixed values of the temperature to determine the dependence of Z_S upon L. Subtraction of each Z_S *versus* L curve from the corresponding stress-strain curve then gave the dependence of $Z_H\simeq Z_U$ upon L.

Figure 6 shows the stress-strain curve obtained for the Butyl rubber at $T=T_0=40$°C, along with its entropy and enthalpy components. The solid lines are

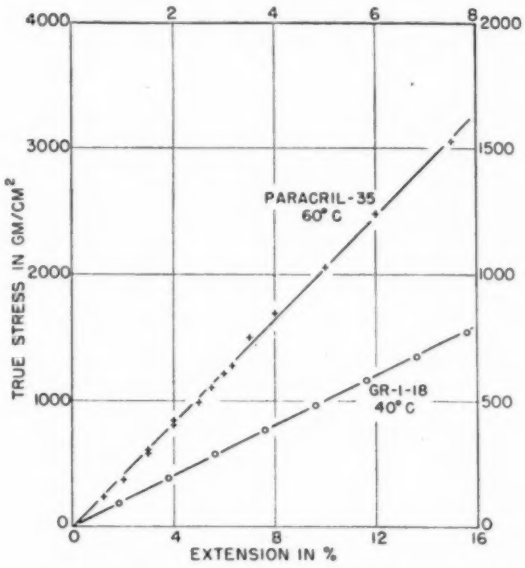


FIG. 5. "True stress"-strain curves. Use scales along top and right-hand margins for GR–I–18. Scales along bottom and left hand margins are for Paracril-35.

the experimental curves, the corresponding curves predicted by the Guth-James theory being shown by dashed lines. Figure 7 presents the same information for the Paracril rubber, at the reference temperature $T_0=60$°C. In general, Figs. 6 and 7 indicate good agreement between the theoretically predicted values of Z, Z_S, and Z_U, and their experimental counterparts. For each rubber, however, the theoretical stress-strain curve tends to deviate above the experimental curve at the high extension end. The discrepancies observed in Fig. 6 between the theoretical and experimental values of Z_S, and $Z_H\simeq Z_U$, are partly the result of experimental error. If the experimental data are internally consistent, then, for the method of fitting described above,[7] the theoretical and experimental values of Z_S and $Z_H\simeq Z_U$ must agree at $L=1$ and $T=T_0$. This conclusion follows immediately from the

fact that for both theory and experiment we must have $E_0 = -3Z_S/\alpha T_0 = -3/\alpha(\partial Z/\partial T)_{LP}$, at $L=1$ and $T=T_0$. Young's modulus computed from this relation should agree with the value determined from the stress-strain curve at the origin. In this instance the value of E_0, computed from the experimentally determined value of Z_S at $L=1$ and $T=T_0$, is 10.4×10^3g/cm², whereas the value obtained from the experimental stress-strain curve was 9.96×10^3g/cm². This error, about 4 percent, probably results from uncertainties in the determination of the point slopes from the experimental stress-temperature curves.

Comparisons between theory and experiment, similar to those shown in Figs. 6 and 7, were also made at several lower temperatures. The same values of K employed earlier were used in these comparisons, since K is assumed to be independent of temperature.

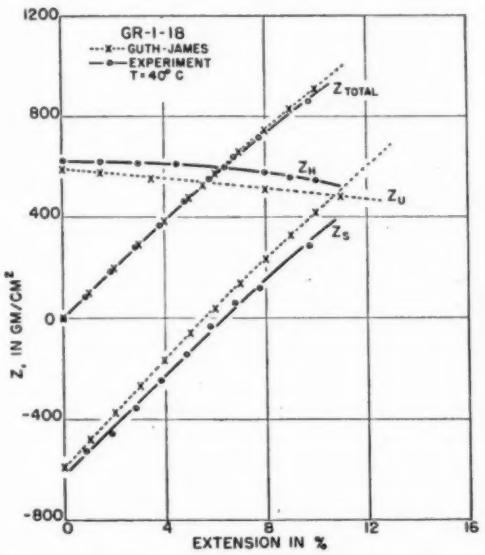


FIG. 6. Theory *versus* experiment for GR–I–18 at 40°C, showing the total stress Z, the enthalpy component $Z_H \simeq Z_U$, and the entropy component Z_s as functions of the percentage extension.

"Experimental" strain values for zero stress were computed with the aid of the linear coefficients of thermal expansion. In each case the theoretical stress-strain curve agreed rather well with its experimental counterpart, but tended to deviate above the experimental curve at the high extension end. The theoretical entropy and internal energy components showed somewhat stronger deviations from the experimental curves. This was particularly true for the Paracril rubber at low extensions and temperatures.

In order to eliminate the influence of uncertainties arising from the evaluation of point slopes, direct comparisons were also made between the theoretical and experimental stress-temperature curves. Figure 8 shows such a comparison for two Butyl curves and two Paracril curves. It appears from Fig. 8 that the temperature dependence of the stress predicted by the theory is a little too strong. If the curves were accurately

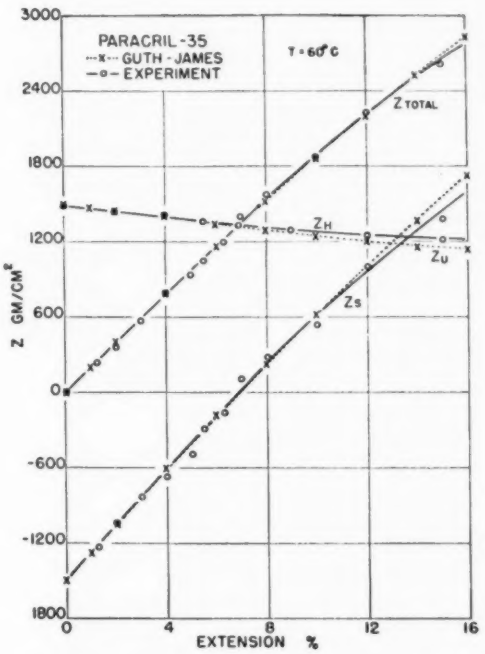


FIG. 7. Theory *versus* experiment for Paracril-35 at 60°C, showing the total stress Z, the enthalpy component $Z_H \simeq Z_U$, and the entropy component Z_s as functions of the percentage extension.

fitted at the high temperature ends, then the theoretically predicted stress values would be too small at the lower temperatures.

Finally, in Fig. 9 the observed dependence of the thermoelastic inversion extension upon temperature is compared with that predicted by the Guth-James theory for the two rubbers. For the Butyl rubber the theoretically predicted values of the thermoelastic inversion extension appear to be somewhat low. About half of this discrepancy, however, can be ascribed

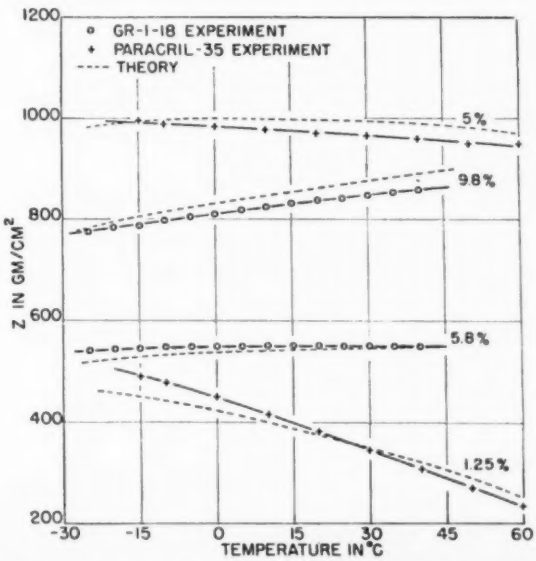


FIG. 8. Theory *versus* observed temperature dependence of stress.

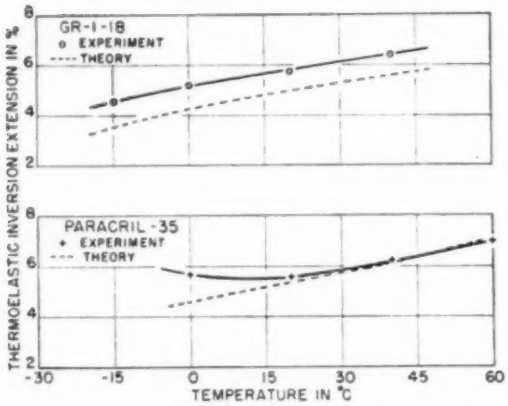


FIG. 9. Dependence of the thermoelastic inversion extension upon temperature.

to the previously noted small inconsistency in the Butyl rubber data. The agreement is very good for the Paracril rubber down to temperatures in the vicinity of +20°C. At lower temperatures the observed thermoelastic inversion extension deviates sharply upward from the predicted values.

ERRORS

The absolute errors in Z, L, Z_S, and Z_U, are difficult to estimate, and no attempt has been made to show them on the graphs. They involve such factors as nonuniformity in specimen thickness, with consequent variations in cross-sectional areas and uncertainties in the stress values based on these areas. Difficulties are experienced in determining the exact value of the specimen length, corresponding to an unstrained condition, when the specimen is mounted in the clamps. This in turn leads to errors in the relative lengths or percentage extensions. In addition, there is the well-known variation in physical properties encountered in going from one rubber specimen to another, even when both specimens are died out of the same tensile sheet. These errors, taken collectively, are best illustrated by the scatter of the experimental points about the true stress-strain curves of Fig. 5.

Percentage extensions are believed to be reliable to better than 0.2 percent extension, and stress values to better than 0.5 percent. The values of Z_S and Z_H are probably reliable to no better than 5 percent since their determination involves the evaluation of point slopes from the experimental stress-temperature curves.

CONCLUSION

The results of this study may be summarized as follows:

(1) The agreement observed between theory and experiment is remarkably good in this region of low extensions—almost to within the estimated experimental errors.

(2) The predicted nonlinear dependence of stress upon temperature is confirmed by the experimental data.

(3) In agreement with the Guth-James theory, the thermoelastic inversion extension is found to be a function of temperature, and is observed to increase with increasing temperature.

(4) Small deviations from the predicted dependence of the stress upon temperature are observed. These appear to increase in magnitude with decreasing temperature, and are strongest for the Paracril rubber at low extensions and temperatures. It seems probable that these deviations are the result of intermolecular forces not specifically included in the theory. That the deviations should be stronger for the Paracril-35 rubber is not at all surprising in view of the highly polar nature of the nitrile groups. At the higher temperatures employed in the experiment such forces may be neglected to a first approximation, because of the greater average distances between molecular groups and the greater thermal kinetic energies. At lower temperatures, however, these forces are no longer negligible, and the molecules exhibit an increasing tendency to "lock-up" as the glassy region is approached.

(5) Finally, an inspection of Figs. 6 and 7 reveals a tendency for the theoretical stress-strain curve to lie above the experimental curve at the higher extensions. Recent studies indicate that this is a real deviation, and that it increases quite rapidly with increasing extension. Similar deviations at somewhat higher extensions have been noted by Treloar,[9] and by Gumbrell, Mullins, and Rivlin.[13]

ACKNOWLEDGMENTS

The authors wish to acknowledge their indebtedness to the Office of Naval Research and to the United States Rubber Company for their financial support of part of this work, and to Dr. E. Guth for many valuable discussions.

[13] Gumbrell, Mullins, and Rivlin, Trans. Faraday Soc. **49**, 1495 (1953).

Thermionic Emission Microscopy of Metals. Part I. General

R. D. HEIDENREICH
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received October 20, 1954)

An electrostatic emission microscope operating with an accelerating voltage of 10–25 kv is briefly described. Factors determining image quality, particularly contrast and resolution, are discussed and the interpretation of images on the basis of the Schottky emission equation is considered. The instrument can be operated at magnifications of 250, 1000, 2500, and 4000× with a resolution consistently better than 1000 A.

It has been found that the iron group metals can be activated easily using barium formate rather than evaporated Ba or Sr. The level of thermionic emission from such surfaces is determined primarily by the partial pressure of chemical reducing agent at the surface (such as carbon) to reduce BaO to free barium. The effect of various impurities and alloying metals is discussed from the standpoint of interpretation of emission images. The most important single agent for promoting emission from Ni, Cu, and Fe, etc., is carbon.

Examples of emission images include nickel, stainless steel, and plain carbon steel. Complications such as surface roughness are demonstrated with micrographs and the limitations of the technique pointed out.

INTRODUCTION

THE structure of electron emitting surfaces is of interest both from the standpoint of the importance of emission characteristics *per se* and from the information that may be gained concerning the structure of metals. The latter is of chief interest here and is a natural extension of studies of thin BaO layers on metal surfaces. The thermionic properties of the underlying metal make themselves evident in emission images when the barium oxide layer is less than about 100 A thick. Consequently, some experiments were undertaken to examine the metallurgical possibilities of the thermionic microscope. Its potentialities have been indicated at various times by results published by Burgers,[1] Rathenau,[2] Gauzit, Septier,[3] and others.

The unique feature of the emission microscope which makes it very attractive for metallurgical research is the ability to observe transformations, crystal growth, etc., at the temperature and time at which they occur. The general usefulness of the technique is dependent upon the correlation of emission with crystallography and chemical composition.

INSTRUMENTAL

The instrument used in this work is entirely electrostatic and is patterned closely after that of Mechlenberg.[4] The projection lens is the double unipotential lens described by Mahl.[5] The magnification of the instrument using the objective lens only is 250× and most of the work was done at this magnification in order to observe larger areas of the emitter. Using the projection lens, magnifications of 1000, 2500, and 4000 are obtained. The higher magnifications were employed for visual examination of fine detail. Since the resolution is practically independent of magnification, it generally proved advantageous to take pictures at 250× and then enlarge photographically. In the range of accelerating voltages employed, 10–25 kv, Ilford Contrast Plates (2 in. ×10 in.) and D-19 developer proved quite satisfactory with exposure times ranging from 2 sec to 20 sec.

The arrangement of the optics is shown in Fig. 1. The emission microscope is actually only part of the instrument, the other portion being an electron diffraction camera. Since the electron diffraction camera is not pertinent to this paper it has been omitted. However, the instrument is constructed so that reflection patterns can be taken from the emitter simply by lifting it away from the objective lens and adjusting to grazing. This feature requires a more elaborate speci-

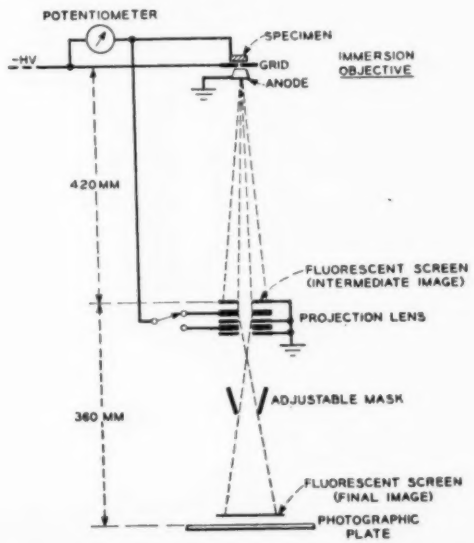


FIG. 1. Schematic of electrostatic emission microscope showing dimensions, location of lenses, and viewing screens.

[1] W. G. Burgers and J. J. A. Ploos van Amstel, (a) Physica **4**, 5 (1937); (b) *ibid.* **4**, 15 (1937); (c) *ibid.* **5**, 305 (1938); and (d) *ibid.* **5**, 313 (1938).
[2] G. W. Rathenau and G. Baas, Physica **17**, 117 (1951).
[3] M. Gauzit and A. Septier, Bull. microscop. appl. **1**, 109 (1951).
[4] W. Mechlenberg, Z. Physik **120**, 21 (1942).
[5] H. Mahl, Metallwirtschaft **19**, 488 (1940).

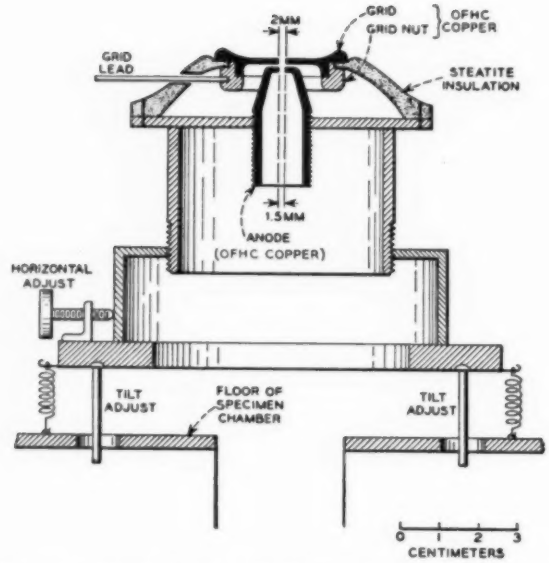


FIG. 2(a). Detail drawing showing mechanical construction of the cathode lens in Fig. 2(b). The spacing of grid and anode can be adjusted as desired.

men manipulator than would be necessary for the microscope alone.

The high tension supply for the microscope is a simple, unregulated rf supply[6] with positive to ground and variable from about 8–28 kv as used in the instrument. Regulation is not so important in an electrostatic instrument. It also turns out that the absence of regulation is an advantage when the objective lens becomes contaminated and breakdown occurs.

The cathode lens is shown in Fig. 2(a). Focusing is accomplished first by mechanically positioning the cathode over the grid and then final focusing done by varying the grid potential. The grid is usually operated about 200 volts negative with respect to the cathode. A 200-megohm high voltage resistor in series with 5-megohm and 10-megohm potentiometers to ground allows both coarse and fine electrical focusing. Small changes in magnification occur during focusing but are not sufficient to be troublesome. The magnification reproduces within 5 percent from one cathode to another.

Figure 2(b) is a schematic of the cathode lens similar to those described in the literature. The shape of the grid aperture is based on early experiments by Johannson[7] and serves to increase the size of the undistorted region of the image. Discussions of this objective will be found in books dealing with electron optics and electron microscopy. Recent investigations of the lens with various modifications have been published by Septier.[8] This lens is quite satisfactory for general emission work, can be easily removed for cleaning, and will operate for several hours before contamination becomes troublesome. The first version of this lens was machined from stainless steel but investigation showed that oxidizing gases from the steel resulted in poisoning of the cathode. The change to OFHC copper with a 1-hr anneal in H_2 at 1000°C before assembly of the lens resulted in greatly improved activities of the various cathodes studied.

The theoretical resolution of the objective is determined by the initial velocities of the emitted electrons and is closely approximated by the simple formula

$$\delta = C(V_0/E), \tag{1}$$

where C is a constant (nearly unity), ϵV_0 is the mean, lateral thermal energy of the emitted electrons and E is the axial accelerating field. V_0 is given by

$$V_0 = (k/\epsilon)T = 8.64\times10^{-5}T \text{ volts}, \tag{2}$$

where k is Boltzmann's constant and ϵ the electron charge. For $T=1160$°K, $V_0=0.10$ volt and for a nominal accelerating field of 5×10^4 volts/cm the theoretical resolution is $\delta\simeq200$ A. Although the instrument is mu-metal shielded and vibration precautions employed, the smallest separation yet measured in an image is about 600 A. Judging from a large number of micrographs, the "useful" resolution obtained in the general run of pictures is about 900 A. This is illustrated in Fig. 3 showing a 900 A separation of lamellae in a steel cathode at 680°C. The cathodes are heated by a bifilar tungsten coil (0.008 in. diam wire) and the heater current supplied by a Variac through an oil immersed filament transformer followed by a full wave selenium rectifier and 60-cycle filter. For minimum distortion in the image, the cathode and lens are aligned so that momentarily switching off the heater current does not produce an image shift. Electrical alignment is achieved by tilt adjustment of the objective until no image shift occurs when the high voltage is momentarily turned off. Figure 4 shows the cathode-heater assembly employed which allows easy removal

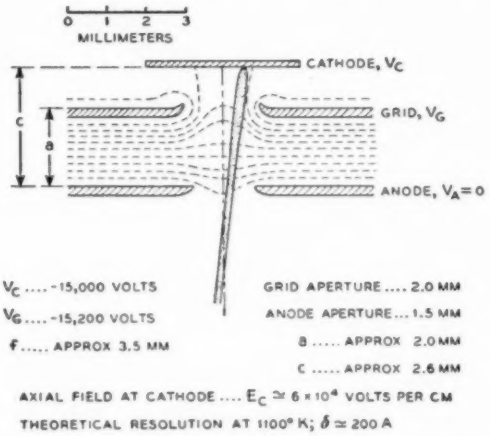
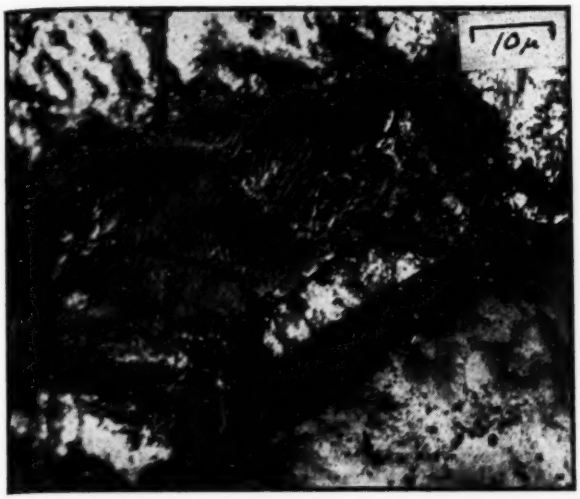

0 1 2 3
MILLIMETERS
CATHODE, V_C
GRID, V_G
ANODE, $V_A=0$
c
a
V_C .... -15,000 VOLTS
V_G .... -15,200 VOLTS
f ..... APPROX 3.5 MM
GRID APERTURE .... 2.0 MM
ANODE APERTURE ... 1.5 MM
a ..... APPROX 2.0 MM
c ..... APPROX 2.6 MM
AXIAL FIELD AT CATHODE .... $E_C \simeq 6\times10^4$ VOLTS PER CM
THEORETICAL RESOLUTION AT 1100° K; $\delta \simeq 200$ A

FIG. 2(b). Electrostatic cathode lens showing shape of grid aperture, equipotentials, and representative electron bundle. The numerical data apply to the illustrations in both Part I and Part II except where higher accelerating voltages are specified.

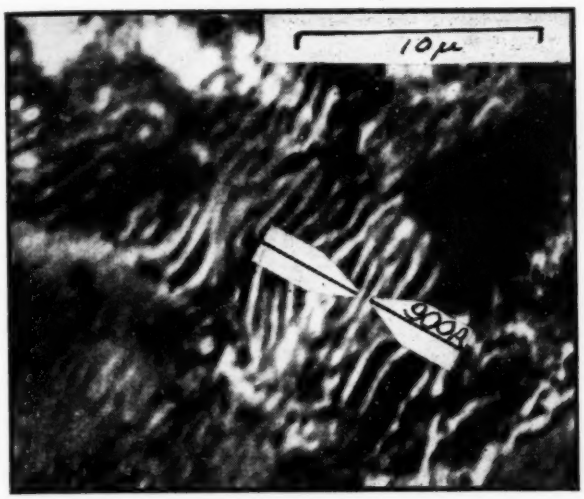
[6] Purchased from Embassy Engineering, 224 East 20th Street, New York City.

[7] H. Johannson, Ann. Phys. **18**, 385 (1933); *ibid.* **21**, 274 (1934).

[8] A. Septier, Compt. rend. **235**, 609 (1952); *ibid.* **235**, 652 (1952); *ibid.* **235**, 1621 (1952).



(a)



(b)

FIG. 3. Emission micrograph illustrating image quality and resolution of fine, lamellar precipitate in pure iron—0.77 percent carbon alloy. Barium activated. Micrograph taken at 250× and photographically enlarged. 700°C, 18 kv.

and replacement of the cathode for electropolishing or chemical treatment of any kind.

Temperatures were measured by moving the cathode by means of the manipulator to a position to the side of the lens where a chromel-alumel thermocouple (0.005-in. wire) was firmly contacted by the cathode surface. Temperatures so measured were checked against a calibration curve obtained with an identical couple spot welded to the cathode and agreed within about 5°C.

INTERPRETATION OF IMAGES[9]

The analysis of thermionic emission data is usually based on the Schottky equation. Preliminary measurements of emission currents to image points using a Faraday collector indicate that the behavior of the surface to be discussed cannot be intelligibly interpreted using the Schottky and Richardson methods. Observations regarding the characteristics of the emission images lead to the conclusion that image contrast and resolution improve as the cathode temperature is lowered and the accelerating potential increased.

In order to obtain useful emission intensities in the temperature range 600–1000°C, it is necessary to apply thin films of electropositive metals such as Ba, Sr, Cs, or Rb. The adsorbed layers produce an outwardly directed positive dipole moment, η, which results in a sizable reduction in work function, $4\pi\eta$. The present knowledge indicates that an activator such as Ba forms a mobile, adsorbed monolayer[10] on the metal surface and that the density of adsorbed atoms varies greatly from one crystal face to another and from one phase to another. This effect is very pronounced as illustrated in Fig. 5 showing a photomicrograph[11] (a) and emission micrograph (b) of the same, identical area on a high purity nickel surface. The crystal faces of different orientation are the dominant image characteristics with surface roughness in evidence in both images. When it becomes possible to correlate the

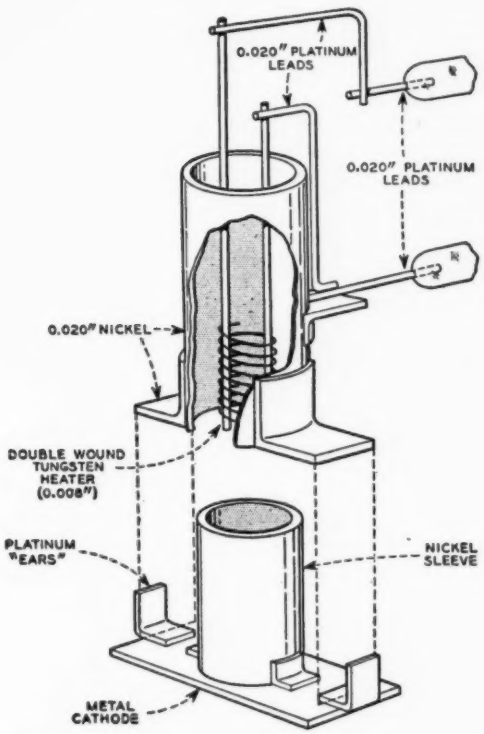
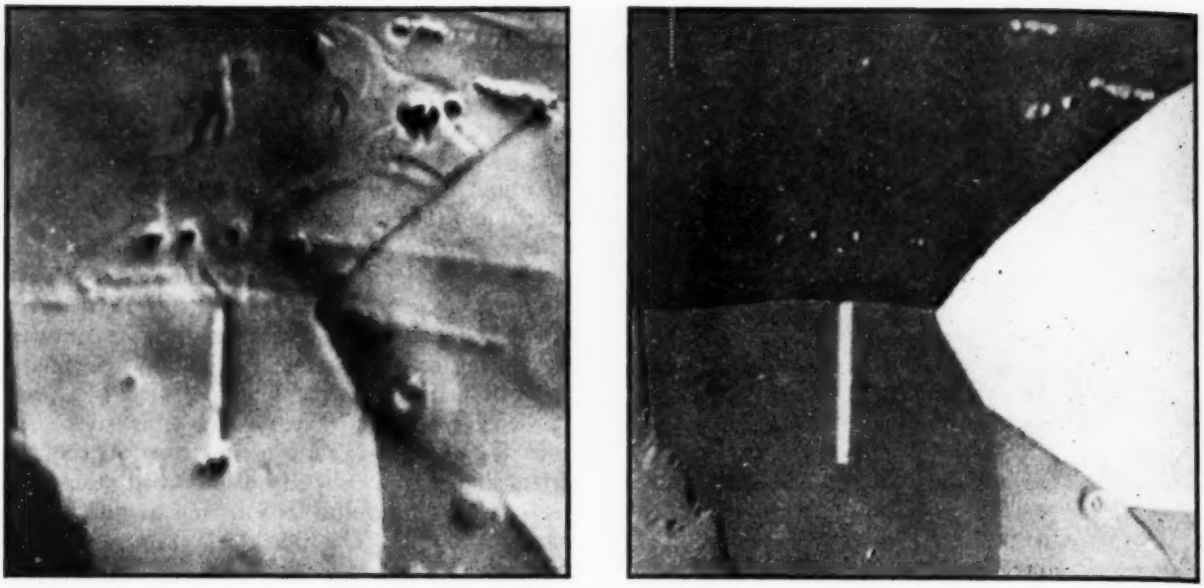


FIG. 4. Construction of heater assembly for demountable cathodes used in the emission microscope.

[9] For a comprehensive discussion of thermionic emission see: C. Herring and M. H. Nichols, Revs. Modern Phys. **21**, 185 (1949).

[10] See J. A. Becker, Trans. Am. Electrochem. Soc. **55**, 153 (1929); J. A. Becker and C. D. Hartman, J. Phys. Chem. **57**, 153 (1953); R. O. Jenkins, Repts. Progr. Phys. **9**, 177 (1943); S. T. Martin, Phys. Rev. **56**, 947 (1939).

[11] The author is indebted to Mr. F. G. Foster of these laboratories for the photomicrograph.



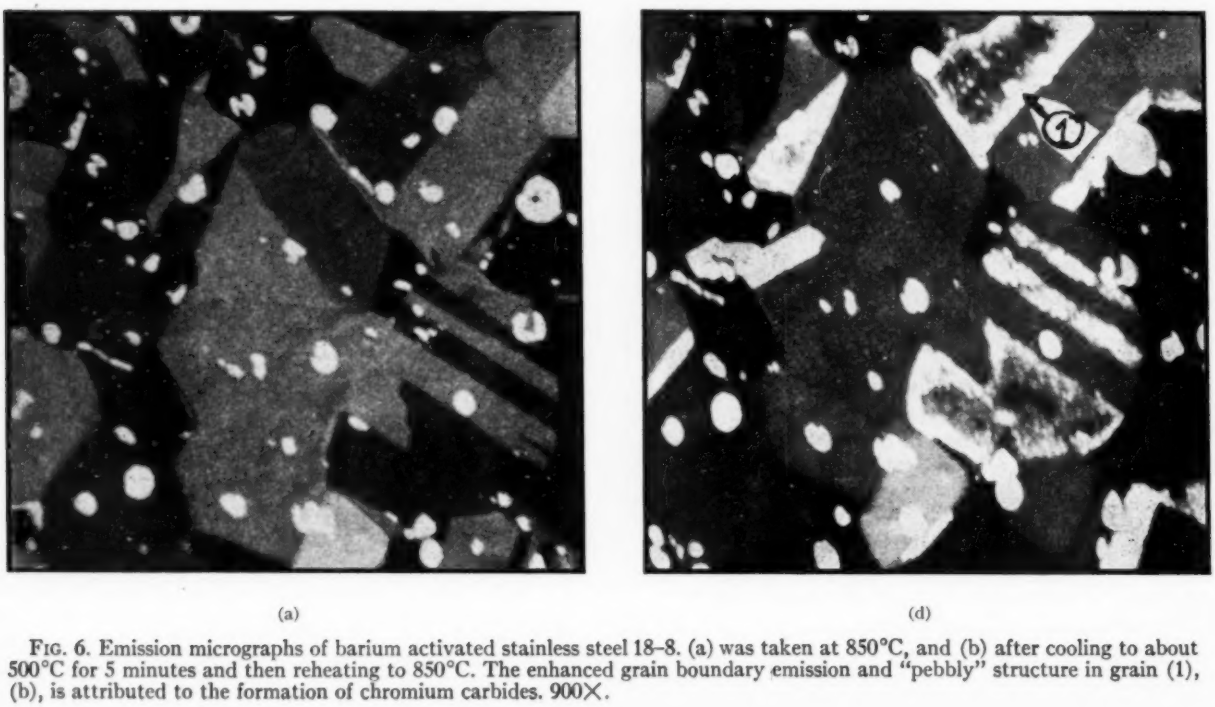
(a) (b)

FIG. 5. Light microscope image (a) and electron emission image (b) of electropolished cathode nickel surface activated with barium. The variation of thermionic work function with crystal face accounts for the contrast between grains in (b). The light faces have low work function. Surface irregularities are also imaged in both (a) and (b). 500×.

emission intensity of a crystal face with the Miller indices of that face, pole figures for a given metal can be rapidly obtained in addition to the grain size now evident in the images.

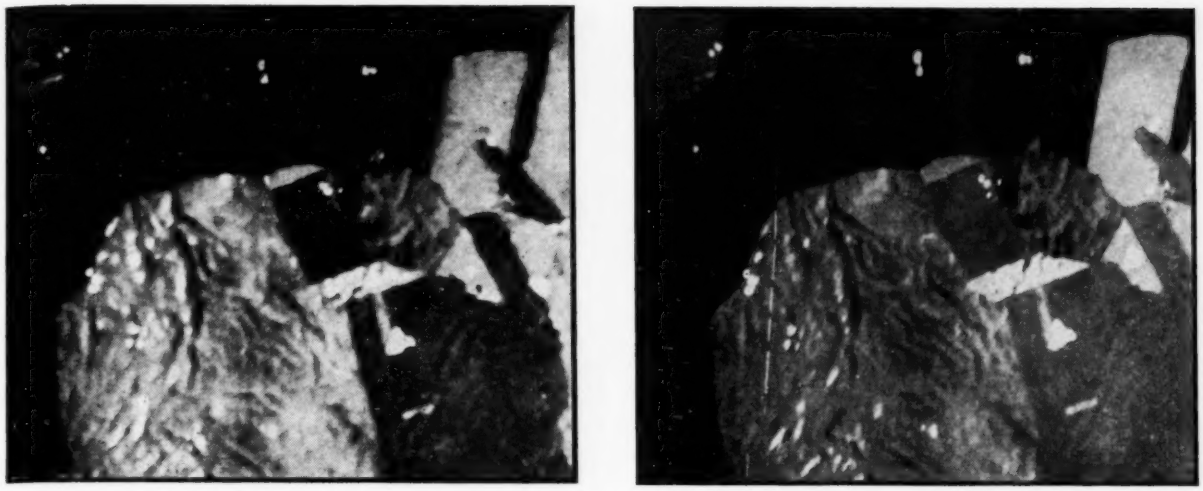
In a given emission image the most pronounced features are usually crystallographic. Changes in chemical composition are often detectable and sometimes quite obvious. In going from one metal or alloy to another very great changes in over-all emission are encountered, and these must be attributed to chemical composition, chiefly at the impurity concentration level.

The two main features of emission images (from smooth surfaces) are demonstrated in Fig. 6 by a sample of 18–8 stainless steel. Figure 6(a) is an emission micrograph taken after annealing at about 850°C and



(a) (d)

FIG. 6. Emission micrographs of barium activated stainless steel 18–8. (a) was taken at 850°C, and (b) after cooling to about 500°C for 5 minutes and then reheating to 850°C. The enhanced grain boundary emission and "pebbly" structure in grain (1), (b), is attributed to the formation of chromium carbides. 900×.



(a) (b)

FIG. 7. Surface roughness or "hill and valley" structure developed on crystal surfaces of austenite in a 0.80 percent carbon steel. (a) Taken at 850°C and (b) at 750°C, showing increased contrast at the lower temperature. Barium activated. 900×.

is chiefly a crystallographic image. After slowly cooling and then reheating, the chemical composition changes become evident. The effect illustrated in Fig. 6(b) is attributed to formation of chromium carbides in the grain boundaries. The large bright spots are due to inclusions or large precipitate particles and are visible in light microscope images as well.

Surface topography often is evident in emission micrographs as mentioned in connection with Fig. 5. Pure metals heated in high vacuum to temperatures of 900–1000°C generally tend to exhibit very smooth crystal surfaces unless the vapor pressure is sufficiently high to produce etching. Alloys, on the other hand, sometimes show a hill and valley effect after heating. This hill and valley effect is much more pronounced on some crystal orientations than on others as illustrated with an SAE–1080 steel at 800°C in Fig. 7(a). Lowering the temperature to 750°C, Fig. 7(b), increases the contrast in the structure as expected. The discussion by Herring[12] in this connection may be pertinent to these observations.

Another effect commonly observed in emission micrographs is that of "ghost" grain boundaries. When a metal is heated in vacuum, surface tension operates to produce cusped grain boundaries or grooves.[13] A subsequent recrystallization or transformation resulting in a new grain structure still shows the previous boundaries until they are thermally smoothed out. Figures 8(a) and 8(b) illustrate this phenomenon. Figure 8(a) shows austenite grains at 800°C in an SAE–1010 steel. The specimen was cooled below the transformation temperature and then raised again to 800°C resulting in a new austenite grain structure. Examination of Fig. 8(b) shows the "ghost" grain boundaries of (a) superposed on the new structure. This effect is often useful since the old boundaries serve as markers for measuring changes in the structure.

[12] Reference 8, Appendix III. Also, C. Herring, Phys. Rev. **82**, 87 (1951).

[13] G. Rathenau, Rev. mét. **48**, 923 (1951). Also reference 2. The more pronounced grooves are associated with high-energy grain boundaries. Low-energy boundaries leave only faint traces while twins generally leave no trace.

CHEMISTRY OF ACTIVATION

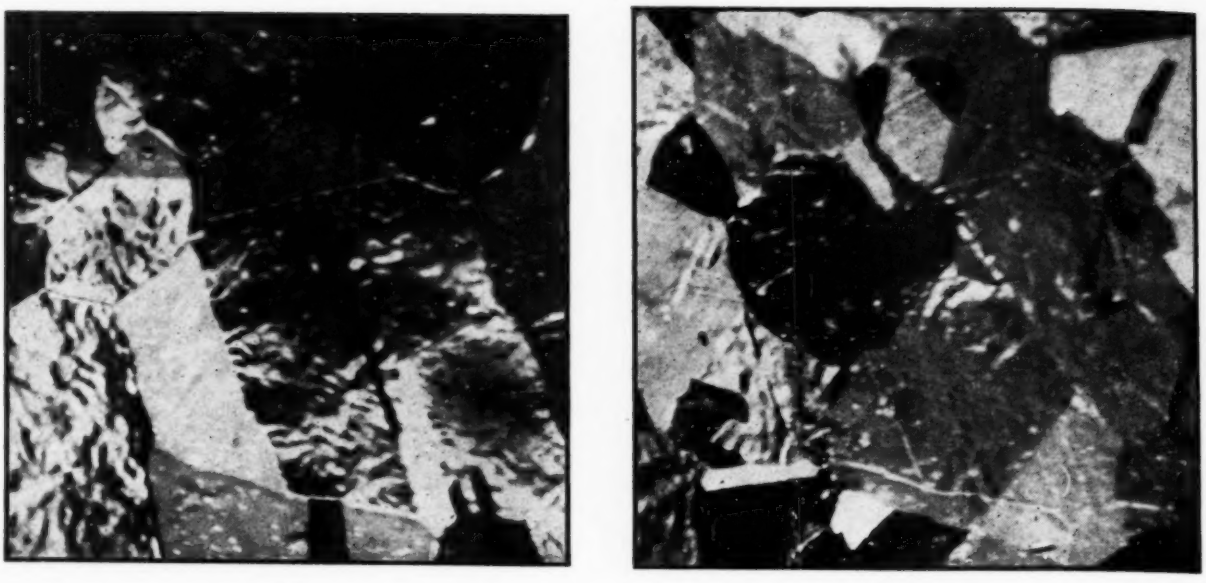
The activation of a metal surface by electropositive atoms such as Ba, Sr, Cs, Rb, and Th has been reported throughout the literature of electron emission and all will work under certain conditions. Barium and strontium are the most practical and reproducible activators found in this work. In this section, the discussion will be confined to barium although the principles apply to the others.

It is concluded from this work and from that of a number of investigators that the fundamental chemical reaction of activation is as follows:

$$\text{(a)}\quad \mathrm{BaO}+\mathrm{R}+\mathrm{O_2}\rightleftarrows\mathrm{Ba}+\mathrm{RO}+\mathrm{O_2}$$

$$\text{(b)}\quad \mathrm{Ba}+\mathrm{M}\rightleftarrows\mathrm{Ba}^{+z}+z\epsilon^{-}\ (z=\text{effective degree of ionization}), \quad (3)$$

where O_2 in (a) represents the residual oxygen in the system. These reactions are written in the simplest form with no regard to balancing in (a) since this is not a pertinent feature. In (3), R denotes a chemical reducing agent and M the metal whose surface is to be activated. These reactions apply whether the activator is originally applied as an evaporated metallic film or as an oxide, carbonate, or other compound. The direction in which (3)(a) tends to go will be determined by the relative partial pressure at the surface of oxidizing and reducing agents. Although activator layers have been applied by evaporation in much of the work reported in the literature and also in this work, it was found that equally good results could be obtained using



(a) (b)

FIG. 8. Emission micrographs of austenite in a 0.80 percent carbon steel (a), original field at 800°C and (b) after transforming rapidly at 600°C and reheating to 800°C to produce new austenite grains. The grooves formed at the grain boundaries of (a) are visible as "ghosts" in (b) and serve as useful markers in following crystal changes. The high-energy boundaries give rise to the most pronounced grooves. 900×.

organic barium compounds applied from solution. The most convenient and best method of application discovered has been barium formate in a solution of 3 parts formic acid and 1 part isopropyl alcohol. The barium concentration in the solution to produce an optimum amount of activator is about 5×10^{-7} moles/cm^3 or the equivalent of about 10 atom layers of barium on the surface. The procedure in preparing a specimen for emission microscopy is then as follows:

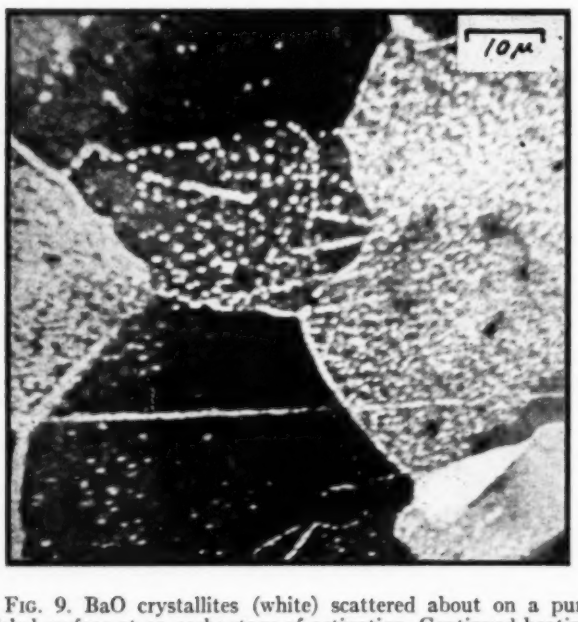


FIG. 9. BaO crystallites (white) scattered about on a pure nickel surface at an early stage of activation. Continued heating produces more free barium at the expense of the BaO and the development of an intense image of the nickel grains.

(1) make up the cathode with heater sleeve as shown in Fig. 4;

(2) abrade surface through 4/0 metallographic paper;

(3) electropolish, etch lightly in $FeCl_3$-HCl in methanol and rinse: methanol→acetone→toluene followed by drying quickly in warm, dry air;

(4) vacuum heat 10–20 min at 900–1000°C;

(5) re-electropolish and rinse as in (3);

(6) apply sufficient activator solution to yield about 10 layers of barium on surface—the solution spreads and dries uniformly on a clean metal surface;

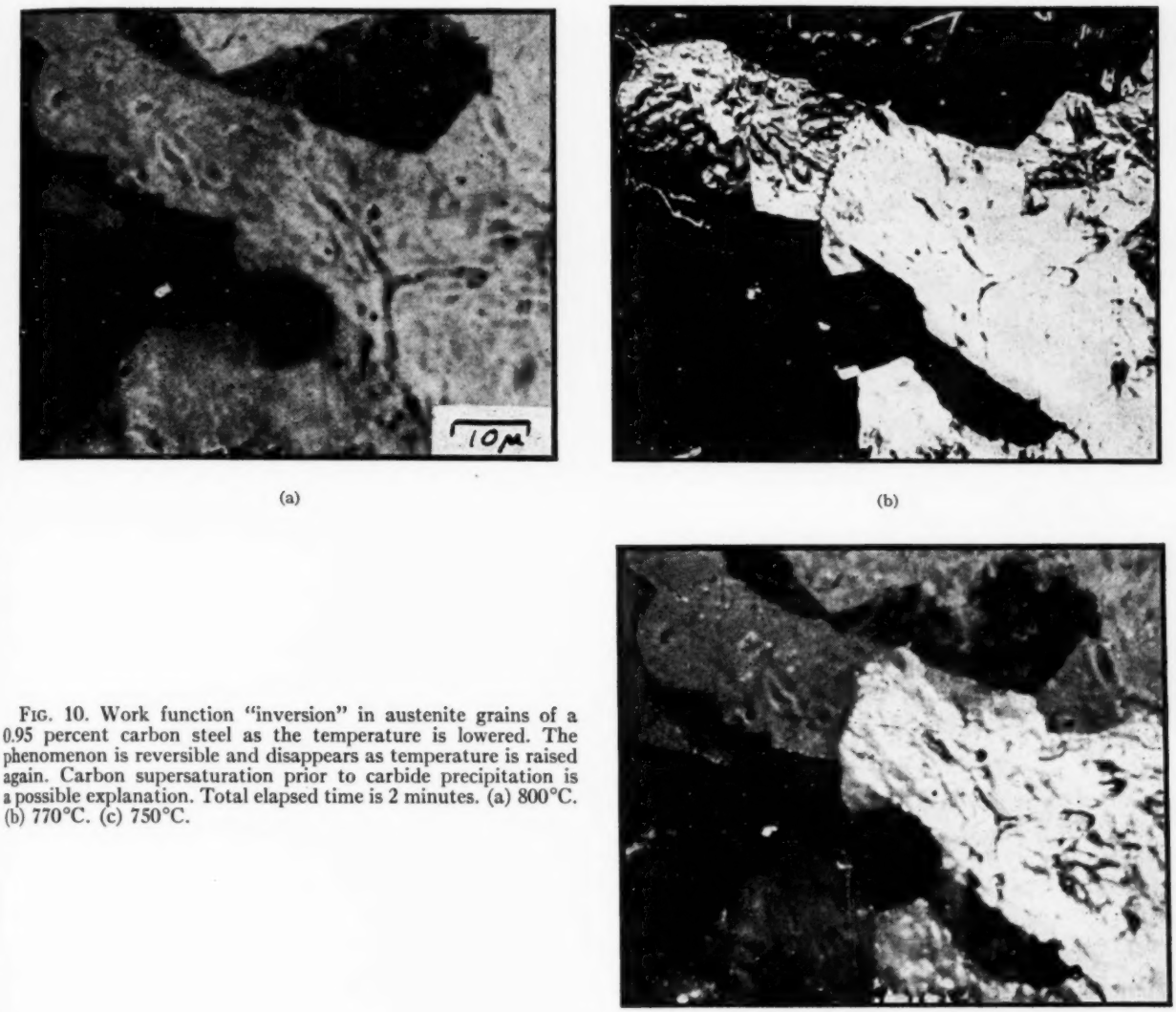
(7) put cathode assembly in microscope—heat cathode to about 800–900°C in period of about 2 minutes and then observe emission.

The temperatures given in this procedure are generally used for iron, steels, nickel, and copper.

Observation during initial breakdown showed that the barium formate breaks down to BaO at a temperature below 400°C probably according to the reaction (unbalanced)

$$Ba(HCO_2)_2(\text{heat/vacuum})\rightarrow BaO+CO+H_2O+C. \quad (4)$$

The carbon produced serves as R in Eq. (3) and so results in an initial high activity *provided* the H_2O has not oxidized the metal surface to form an oxide which is not reduced again at the temperature of observation. If the cathode metal oxide is easily reducible to metal, Eqs. (4) and (3) will operate to yield emission at temperatures as low as 650°C. This emission will be only transient, however, if the residual vacuum and cathode itself do not maintain a sufficient partial pressure of



(a) (b) (c)

FIG. 10. Work function "inversion" in austenite grains of a 0.95 percent carbon steel as the temperature is lowered. The phenomenon is reversible and disappears as temperature is raised again. Carbon supersaturation prior to carbide precipitation is a possible explanation. Total elapsed time is 2 minutes. (a) 800°C. (b) 770°C. (c) 750°C.

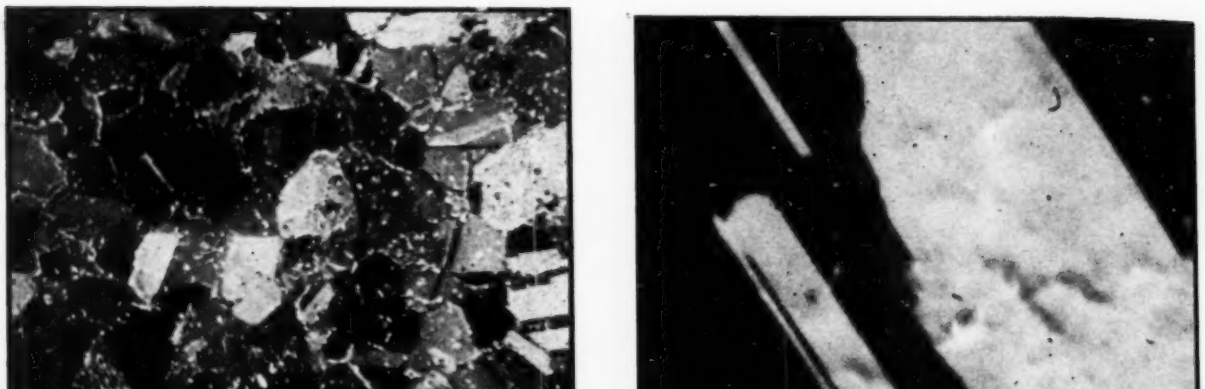
reducing agent, R, so that Eq. (3) can be constantly driven to the right.

Figure 9 illustrates the emission from high purity nickel shortly after breakdown [Eq. (4)]. BaO crystals with high emission are seen scattered about on the surface and the grain structure becoming evident as reaction (3) proceeds. In general, BaO crystals a few hundred angstroms in size exhibit lower work function than the adsorbed ion layer on the metal and appear bright in the image.

The development of grain structure from scattered BaO crystals has been observed a number of times and is of some interest. The adsorbed ion layer starts at the base of the BaO crystal and moves radially away from the crystal which itself shrinks in size as the grain structure of the metal becomes brighter. Free barium metal forms at the BaO crystal by reaction (3)(a). Upon contact with the *clean* base metal M, the barium atoms become at least partially ionized to form barium positive ions. The mutual repulsion of the Ba^{+z} ions is then responsible for the lateral spread as a sort of two-dimensional gas. Different crystal faces accommodate different populations of positive ions or bring about different degrees of ionization. In either case, the result is a positive layer with reduction of work function and development of an emission image occurs.

The most satisfactory reducing agent, R, from the standpoint of emission microscopy is carbon. Ti, Mg, and Si also reduce BaO, but the reaction product is a solid oxide (such as MgO) which is deposited on the metal surface and may interfere with emission.

A rather rough survey of various base metals (as pure as could be obtained) was made from the standpoint of metallurgical studies in the temperature range 650–1000°C. They were classed as good or poor depending upon usable intensity at 250× below 1000°C when activated by the formate procedure. Table I summarizes the findings. This is not a hard and fast classification and some metals listed as poor may well be shifted to the other column by appropriate treatment.



(a) (b)

FIG. 11. Emission micrographs of electrodeposited nickel (a) and austenite (b) in a 0.77 percent carbon steel showing the contrast in emission from twin crystals.

From the standpoint of metallurgy, they are not as interesting at present as are the metals in the good column.

A very important factor is the impurity content of the metals, particularly O_2, N_2, C, and S as well as metallic impurities or alloying elements that may tend to concentrate at the free surface upon vacuum heating. From additions of various metallic elements to the formate activator applied to iron and steel base cathodes, Table II was constructed. Cobalt is very effective in promoting emission from iron and steel and reduces the minimum operating temperature by nearly 100°C. Micrographs of high purity iron could not be obtained without addition of cobalt formate to the activator. The amount of cobalt added is slightly less than the barium concentration.

Table I emphasizes the fact that formate activation is most successful on metals whose oxides are reduced readily at temperatures below 1000°C. For the metals in the "Poor" column, activation by evaporation of barium onto the surface from a separate source in high vacuum is preferable. The oxides of the metals in the "Poor" column are reduced at higher temperatures and then with difficulty.

TABLE I. Emission from base metals below 1000°C. (Ba formate activator.)

Good	Poor
Fe	Ti
Ni	Si
Cu	Zr
Cr	W
Co	Ta
	Mo

The single impurity added to the base metals which gives the most spectacular increase in emission is carbon. The presence of carbon does not guarantee high emission, however, for other impurities may simultaneously behave as poisons. For example, high purity iron containing only carbon from 0.1–0.8 percent gives excellent emission down to 650°C. Commercial steels in the same carbon range cannot be readily observed below 750°C. Interestingly enough, nitrogen has a very pronounced effect and a deleterious one. If a small amount of nitrogen is present in a pure iron-carbon alloy, high emission can be obtained but the surface details "freeze-in." By "freeze-in" it is meant that the initially mobile activator layer becomes immobile and no longer follows crystallographic changes in the underlying metal. Treatment of a very active iron-carbon alloy for only 2 minutes in NH_3 gas at 600°C serves to poison the emission. The mechanism of the "freeze-in" in the one case and the complete loss of emission in the other is not understood. Electron diffraction patterns in both instances indicate the presence of iron nitrides or carbo-nitrides. The freeze-in case with high emission yielded $\epsilon' - Fe_3N$ type patterns

TABLE II. Effect of metallic additives to barium formate activator on iron and steel (below 1000°C).

Promoters	Poisons	Little effect
Co[a]	Ag	Li
Cu	Hg	Na
Mn	Be	K
Cr	Sn	
	Si	
	Ti	

[a] Most effective.

which was probably a carbo-nitride.[14] Nitrogen, then, can be a serious complication in the study of iron-carbon alloys.

Micrographs of all the metals examined are not displayed for reasons of space. OFHC copper gives high emission in the range 700–800°C but the contrast between crystal faces is very low. Above 800°C, thermal etching results and no surface detail can be imaged during evaporation of copper.[15] From a practical viewpoint, emission microscopy above 850°C is more satisfactory using strontium or a strontium-barium activator with cobalt due to the lower vapor pressure of strontium. Barium gives higher emission currents and is superior for low-temperature work. Rb is satisfactory but is evaporated at elevated temperatures. Cesium suffers from the disadvantage that its oxide is very volatile so that it is lost by evaporation. Cesium generally leads to early electric breakdown and grid emission in the objective lens. All of the micrographs exhibited in this paper were obtained by (Ba, Co) formate activation.

There are a number of interesting effects not mentioned as yet. One of these is what might be termed "work function inversion." It occurs in steels near a transformation or precipitation temperature and is illustrated in Fig. 10. As the temperature is lowered certain areas begin to emit electrons with a higher intensity indicating that the decrease in work function more than compensates the decrease in temperature. As the temperature drops these areas become more numerous as seen in Fig. 10. The reason for this is not clear.

One of the more spectacular features in emission images of metals is the high contrast in twins. This is illustrated in Fig. 11 for fine grain, high-purity nickel and for an austenitic steel. Twin bands generally represent the extremes in work function (contrast) seen in an emission image.

As near as can be determined, the mobile activator layer follows underlying crystallographic changes nearly instantaneously. It goes without saying that the value of emission microscopy in metallurgy lies in studying emission images *during* crystallographic and precipitation changes and not in isolated pictures. The predominant features in the image are crystallographic and variations in chemical composition. Consequently the obvious applications include:

(1) Grain size and grain growth studies at various temperatures [reference (2)].

(2) High-temperature transformations (reference 1 and Part II of this paper).

(3) Precipitation and segregation in high-temperature alloys—an example is shown in Fig. 12 which exhibits grain boundary precipitation as well as scattered precipitate in Inconel-X.

The lowest temperature at which the author has been able to maintain useful emission is 630°C. Increasing the accelerating potential is limited since breakdown and grid emission in the objective lens became troublesome.

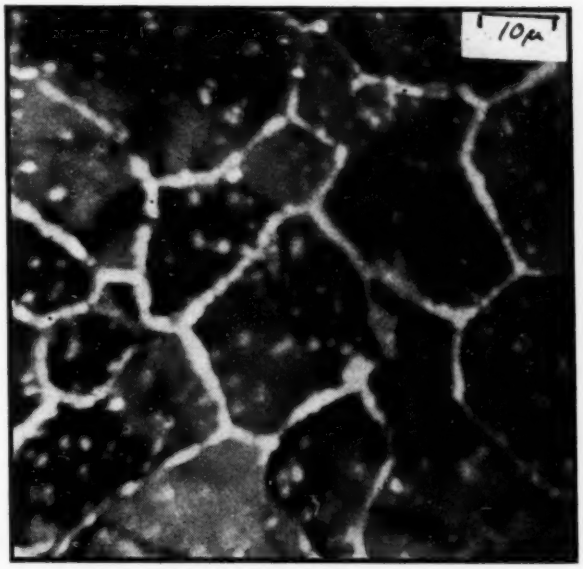


FIG. 12. Emission micrograph of Inconel-X at 850°C showing enhanced grain boundary emission and scattered bright spots probably due to a precipitate in the alloy.

ACKNOWLEDGMENTS

The author is indebted to various members of the technical staff, particularly Mr. C. Herring and Mr. G. E. Moore for aid and criticism in this work.

[14] See K. H. Jack, Proc. Roy. Soc. (London) **A195**, 34 (1948).
[15] Oxygen may be responsible for the thermal etch.

Numerical Solutions of Spherical Blast Waves

HAROLD L. BRODE
The RAND Corporation, Santa Monica, California
(Received December 6, 1954)

The strong-shock, point-source solution and spherical isothermal distributions were used as initial conditions for a numerical integration of the differential equations of gas motion in Lagrangean form. The von Neumann-Richtmyer artificial viscosity was employed to avoid shock discontinuities. The solutions were carried from two thousand atmospheres to less than one-tenth atmospheres peak overpressure. Results include overpressure, density, particle velocity, and position as functions of time and space. The dynamic pressure, the positive and negative impulses of both dynamic pressure and static overpressure, positive and negative durations of pressure and velocity, and shock values of all quantities are also described for various times and radial distances. Analytical approximations to the numerical results are provided.

I. INTRODUCTION

THE problem of a spherical blast in air has been solved analytically for strong shocks in an ideal gas by J. von Neumann,[1] and in the weak shock approximation by H. A. Bethe.[1] For intermediate regions it has been found necessary to resort to numerical methods. The availability of high-speed computing machines has made possible the solution of this problem in the intermediate range with considerable accuracy.

Two other groups are currently engaged in performing this work by independent methods. Under J. von Neumann and H. Goldstine at Princeton a shock fitting method has been employed. The differencing method of Peter Lax[2] is being used at New York University by S. Lowell.

The method used in this paper is due to von Neumann and Richtmyer[3] and employs an artificial viscosity as a mechanism for avoiding shock-front discontinuities. Previously, T. S. Walton reported some results[4] using this method and an initial isobaric sphere of about 13 atmospheres.

The integrating process consists of the step-wise solution of difference equations which approximate the differential equations of motion of the gas. In order for such a procedure to be workable, however, a number of practical conditions must be satisfied. The differencing scheme must be stable, must offer reasonably detailed results, must conserve numerical significance, and when put in the form of coded instructions for a high-speed computer, must be fast enough to reach desired solutions with a reasonable expenditure of machine time.

II. METHOD FOR NUMERICAL INTEGRATION

Equations of Motion

The Lagrangean equations of motion are reduced to dimensionless parameters wherein pressure (P), density (ρ), and velocity (u) are measured in units of ambient pressure (P_0), density (ρ_0), and sound velocity (C_0), respectively. In the following discussion, the pressure will frequently be expressed in atmospheres where one atmosphere is defined as equal to the pre-shock ambient pressure (P_0). Where the expression *overpressure* (ΔP) is used the reference is to the pressure in atmospheres in excess of the ambient pressure (P_0), i.e., $\Delta P = P - 1$. We shall continue to speak of a velocity (u), however, when more properly we might refer to a *mach number* (u). The radial distance $r(r_0,t)$, is expressed in energy-reduced dimensionless units (r_0 being Lagrangean distance, and t the time), such that

$$\lambda = r/\epsilon \quad \text{and} \quad \lambda_0 = r_0/\epsilon, \tag{1}$$

where ϵ is a length expressing the energy and ambient pressure scaling:

$$\epsilon^3 = \frac{E_{\text{tot}}}{P_0} = \frac{4\pi}{P_0}\int_0^R \rho\left(E_{\text{int}} + \frac{u^2}{2}\right) r^2 dr - \frac{4\pi R^3}{3(\gamma-1)}. \tag{2}$$

E_{tot} is the total blast energy and E_{int} is the specific internal energy. The subtracted term represents the pre-shock internal energy of the gas engulfed by the shock, and R is the shock radius. Time (t) is defined in dimensionless units (τ) such that

$$\tau = tc_0/\epsilon.$$

The artificial viscosity (q), which acts like a pressure, is in units of the pre-shock ambient pressure (P_0).

In these units the Lagrangean equations of motion are written as follows:

$$\frac{\partial \lambda}{\partial x} = \frac{1}{\rho\lambda^2} \quad \text{or} \quad \frac{\partial \rho}{\partial \tau} = -\rho\left(\frac{2u}{\lambda} + \frac{\partial u/\partial x}{\partial \lambda/\partial x}\right) \quad \text{(mass)}, \tag{3}$$

$$\frac{\partial u}{\partial \tau} = \frac{-\lambda^2}{\gamma}\frac{\partial}{\partial x}(P+q) \quad \text{(momentum)}, \tag{4}$$

$$\frac{\partial P}{\partial \tau} = -\frac{1}{\rho}\frac{\partial \rho}{\partial \tau}[\gamma P + (\gamma-1)q] \quad \text{(energy)}, \tag{5}$$

$$u = \frac{\partial \lambda}{\partial \tau}. \tag{6}$$

[1] *Shock Hydrodynamics and Blast Waves* (Los Alamos Scientific Laboratory, 1944), Report AECD-2860.
[2] P. D. Lax, Commun. Pure Appl. Math. (Institute of Mathematical Sciences, New York University) **VII**, 159–193 (1954).
[3] J. von Neumann and R. D. Richtmyer, J. Appl. Phys. **21**, 232 (1950).
[4] T. S. Walton, Phys. Rev. **87**, 910(A) (1952).

In these equations the Lagrangean variable is taken to be $x=\frac{1}{3}(r_0/\epsilon)^3$.

In Eq. 5 (energy conservation) the internal energy for an ideal gas has been used.

$$E=\frac{P}{\rho(\gamma-1)}\frac{\rho_0}{P_0}. \tag{7}$$

Form of Artificial Viscosity

An appropriate viscosity for the case of an outward moving spherical shock wave is the following:

$$q=\frac{9\gamma(\gamma+1)}{4}\left(\frac{M}{3\pi}\right)^2\rho(\Delta x)^2\left(\frac{\partial u}{\partial x}\right)\left(\frac{\partial u}{\partial x}-\left|\frac{\partial u}{\partial x}\right|\right) \tag{8}$$

where Δx is the grid size and M is the number of grid zones in the shock front. The form of q chosen here is basically similar to that of von Neumann and Richtmyer, and satisfies similar conditions for the spherical case. However, there is no convenient steady-state solution by which to demonstrate the connection of variables across the shock, and the form chosen here is only asymptotically the same as a verified form for a plane problem.

The fact that the plane form approximation is good for shock thicknesses which are small relative to the shock radius is borne out by the close agreement with a similarity solution (with q) investigated by R. Latter[5] in which he solves numerically the resulting ordinary differential equations for cases of interest to this problem.

The particular form chosen for the viscosity has the advantage that it contributes nothing in regions of expansion, and is nonzero only in the compression phase of a shock. In Lagrangean coordinates this has the advantage of eliminating a spurious contribution near the origin where the positive velocity gradient is large.

Initial Conditions

Two general types of initial conditions were taken, (1) a point source, and (2) an isothermal sphere. A point-source case was run in the greatest length using the von Neumann strong-shock solution beginning at 1600 atmos shock overpressure, and running down to less than 0.06 atmos. The point-source solution, starting at 199 atmos peak overpressure, was used to start the one problem carried out on the ORDVAC machine at the Ballistics Research Laboratories, Aberdeen, Maryland. The latter problem was run to nearly 0.1 atmos shock overpressure. The point-source solution was also used for starts at 473 and 818 atmos shock overpressure, and these problems were run down to less than 100 atmos peak overpressure.

Three isothermal-sphere type problems have been run so far. Two of these begin with a hot isothermal sphere for which the density is the same inside and out. One began at 2002 atmos, the other at 121 atmos, overpressure.

One cold isothermal sphere was also used, beginning with 121 atmos overpressure and a temperature inside equal to the outside temperature.

These three problems were carried down to shock overpressures less than unity.

Stability and Problem Checks

Short plane-geometry problems were found useful in checking the validity of various computational tricks, and in studying the effect of certain changes in the difference equations. The effect of the size of the viscosity constant was also more easily studied in the plane case. Time increments smaller than that prescribed by the *Courant condition* ($\Delta t=(p\rho)^{-\frac{1}{2}}\Delta r$) by a factor of 2 to 4 were found necessary for smooth results. In an attempt to reduce the required computing time the stable differencing scheme of Du Fort and Frankel[6] for diffusion-type equations was employed in place of explicitly carrying a viscosity quantity (q). Unfortunately such a scheme has some practical disadvantages. It requires carrying throughout a machine calculation sets of data for all space points for two different times. Furthermore, computing, changing time increments, and combining space points all become more tedious. Besides these disadvantages, additional terms must be introduced to correct for the excess energy introduced by the differencing scheme. On the other hand, the very general nature of the viscosity method, the ease of its applicability, and the precision with which it reproduces the Hugoniot conditions across a shock would seem to offset the more stringent time requirement. Use of this method for nonideal gases is not considered in this paper, however.

It is frequently convenient to use unequal zone sizes. For instance, the use of small zones through the shock front provides a sharp shock at very little cost in computing time. The use of such unequal zones was empirically validated in this problem by repeating calculations with quite different zone choices.

The size of the time increments were automatically doubled whenever the stability conditions would allow it. Two conditions exist, one being the usual Courant condition and the other a diffusion-type condition imposed by the artificial viscosity

$$\begin{aligned}\Delta\tau &\leqslant \Delta x/\lambda^2(p\rho)^{\frac{1}{2}}{}_{\max}\\ \Delta\tau &\leqslant \frac{\gamma}{4}(\Delta x)^2\left[\frac{1}{\lambda^2 q}\left|\frac{\partial u}{\partial x}\right|\right]_{\min}.\end{aligned} \tag{9}$$

The total energy in the blast wave must be conserved, but a check on the total energy is not a very sensitive test of the correctness of the results. Several machine

[5] R. Latter, J. Appl. Phys. (to be published 1955).

[6] E. C. Du Fort and S. P. Frankel, *Mathematical Tables and Other Aids to Computation*, Vol. VII, p. 135 (1953).

errors were detected (and eliminated) which did not substantially effect the total energy. Here the system employed was to back off and re-run whenever suspicious aberrations occurred and to re-run and compare the entire problem with a different zone spacing, different viscosity, different time increments, as well as some differences in the difference equations.

Difference Equations

The differential equations are approximated by the following difference equations:

$$u_l^{n+\frac{1}{2}}=u_l^{n-\frac{1}{2}}-\frac{\Delta\tau(\lambda_l^n)^2}{(\Delta x)_l\gamma}\left[P_{l+\frac{1}{2}}^n-P_{l-\frac{1}{2}}^n+q_{l+\frac{1}{2}}^{n-\frac{1}{2}}-q_{l-\frac{1}{2}}^{n-\frac{1}{2}}\right], \quad (10)$$

$$\lambda_l^{n+1}=\lambda_l^n+u_l^{n+\frac{1}{2}}\Delta\tau, \quad (11)$$

$$\rho_{l-\frac{1}{2}}^{n+1}=\rho_{l-\frac{1}{2}}^n\left(\frac{1-W}{1+W}\right), \quad (12)$$

where

$$W\equiv\Delta\tau\left(\frac{2(u_l^{n+\frac{1}{2}}+u_{l-1}^{n+\frac{1}{2}})}{\lambda_l^{n+1}+\lambda_l^n+\lambda_{l-1}^{n+1}+\lambda_{l-1}^n}+\frac{u_l^{n+\frac{1}{2}}-u_{l-1}^{n+\frac{1}{2}}}{\lambda_l^{n+1}+\lambda_l^n-\lambda_{l-1}^{n+1}-\lambda_{l-1}^n}\right), \quad (13)$$

$$q_{l-\frac{1}{2}}^{n+\frac{1}{2}}=9\frac{\gamma(\gamma+1)}{2}\left(\frac{M}{3\pi}\right)^3\rho_{l-\frac{1}{2}}^{n+1}\left[u_{l-1}^{n+\frac{1}{2}}-u_l^{n+\frac{1}{2}}\right]^2, \quad \text{for} \quad u_{l-1}^{n+\frac{1}{2}}>u_l^{n+\frac{1}{2}}, \quad (14)$$

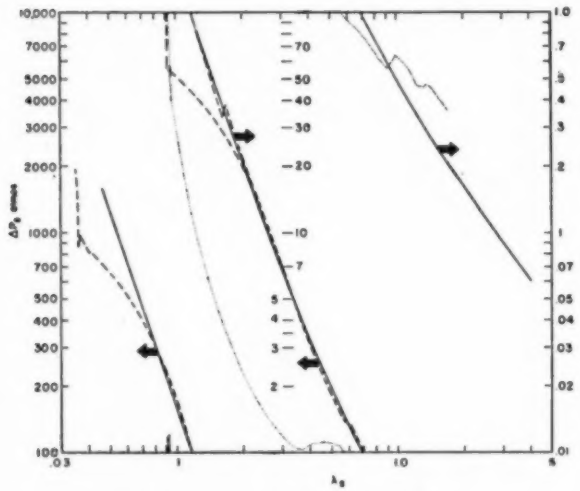


FIG. 1. Radial dependence of the peak or shock overpressure. The solid curve represents the point-source solution. The dashed curves represent the results from two different initial isothermal spheres, one at 2000 atmos and the other at 121 atmos, both spheres were hot with equal density inside and out. The dotted curve represents the resulting shock overpressure from an initially cold isobaric sphere with temperature equal inside and out. The pressures are in atmos and the distances in units of $(E_{tot}/P_0)^{\frac{1}{3}}$.

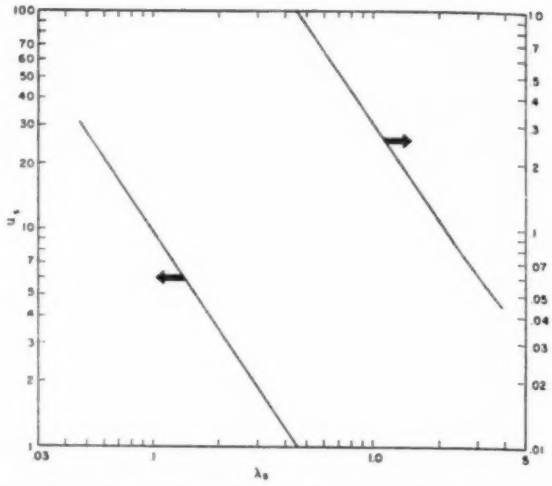


FIG. 2. Particle velocity or Mach number at the shock as a function of shock radius for the point-source solution. Velocity is in units of the pre-shock sonic velocity and the radius is in units of $(E_{tot}/P_0)^{\frac{1}{3}}$.

$$q_{l-\frac{1}{2}}^{n+\frac{1}{2}}=0 \quad \text{for} \quad u_{l-1}^{n+\frac{1}{2}}\leqslant u_l^{n+\frac{1}{2}},$$

$$P_{l-\frac{1}{2}}^{n+1}=\frac{\left[\frac{\gamma+1}{\gamma-1}\rho_{l-\frac{1}{2}}^{n+1}-\rho_{l-\frac{1}{2}}^n\right]P_{l-\frac{1}{2}}^n+2(\rho_{l-\frac{1}{2}}^{n+1}-\rho_{l-\frac{1}{2}}^n)q_{l-\frac{1}{2}}^{n+\frac{1}{2}}}{\frac{\gamma+1}{\gamma-1}\rho_{l-\frac{1}{2}}^n-\rho_{l-\frac{1}{2}}^{n+1}}. \quad (15)$$

III. RESULTS OF NUMERICAL SOLUTIONS

Peak Overpressure (P_s) *versus* Shock Radius (λ_s)

The strong-shock point-source solution provides that the shock overpressure (ΔP_s) should depend on the inverse cube of the shock radius (λ_s). The problems that were begun with strong-shock, point-source values continued to obey the inverse cube law down to 10 atmos at which point the overpressure had become 3 percent higher than the strong shock prediction. The addition of 1 atmos to an inverse cube term gives a relation valid over a greater range of overpressures, deviating from the calculated curve by less than 5 percent at 5 atmos.

$$\Delta P_s=0.1567\lambda_s^{-3}+1 \text{ atmos.} \quad (16)$$

The shock radius (λ_s) is in the dimensionless units (energy/pressure-reduced) described in Sec. II.

At still lower pressures the following empirical fit applies.

$$\Delta P_s=\frac{0.137}{\lambda_s^3}+\frac{0.119}{\lambda_s^2}+\frac{0.269}{\lambda_s}-0.019 \text{ atmos,}$$

$$\text{for} \quad 0.1<\Delta P_s<10, \quad \text{or} \quad 0.26<\lambda_s<2.8. \quad (17)$$

For all the calculations begun with a strong-shock the solid curve in Fig. 1 presents the peak overpressure as a function of shock radius.

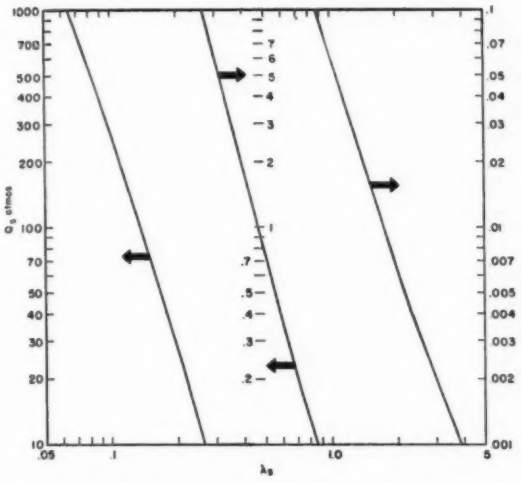


FIG. 3. Peak or shock dynamic pressure ($Q=\frac{1}{2}\rho u^2$) *versus* shock radius for the point-source solution. The dynamic pressure is in atmos and the radius in units of $(E_{tot}/P_0)^{\frac{1}{3}}$.

The dashed curves represent the peak overpressures resulting from the hot isothermal spheres with normal density inside (beginning at 2002 and 122 atmos overpressure). Notice that the overpressure becomes indistinguishable from that of the strong shock at a radius where the mass of air engulfed by the shock is 10 times the initial mass. In the strong-shock region this corresponds to a drop in overpressure to about 17 percent of the initial value, and in any case to a radius a little more than twice the initial radius.

The dotted curve represents the overpressure from an initially cold isothermal sphere (normal temperature, high pressure). Because of the slower rarefaction speed the shock pressure does not rise to the point-source value until quite late, i.e., until the rarefaction which moves inward from the initial pressure front has reached the center.

In each case of an initial isothermal sphere there are some slight oscillations about the point-source overpressure curve which are caused both by rarefactions and small shocks which form behind the front shock, move inward and reflect off the origin and then move out to overtake the shock front. (See a later discussion of these special features of the isothermal sphere problems.)

Peak Dynamic Pressure (Q_s) and Peak Particle Velocity (u_s) *versus* Shock Radius (λ_s)

The particle velocity and the peak pressure at the shock front follow precisely the Hugoniot relation,

$$u_s=\frac{5\Delta P_s}{(49+42\Delta P_s)^{\frac{1}{2}}}. \tag{18}$$

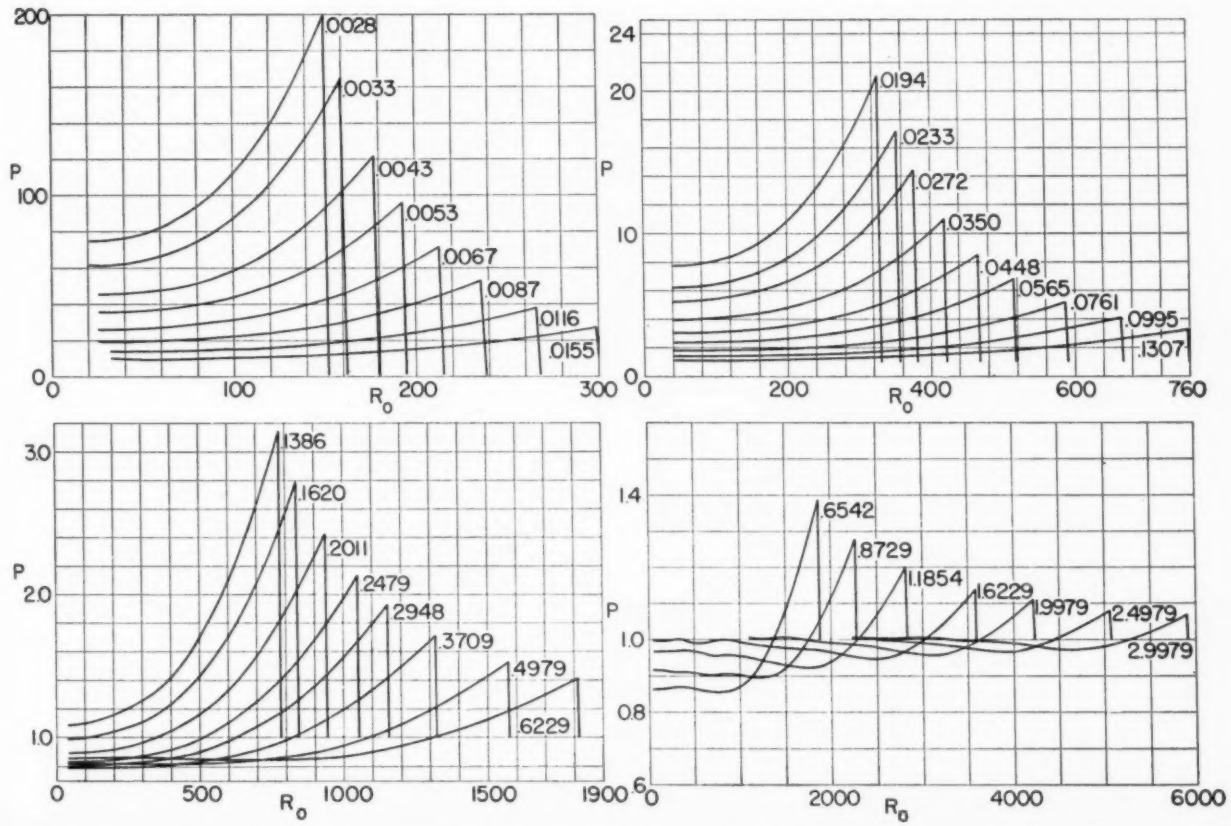


FIG. 4. Pressure in atmos as a function of the Lagrange or mass position for the point-source solution at times indicated. The position is in arbitrary units of $(E_{tot}/P_0)^{\frac{1}{3}}/1627.2$, and the time is in units of $(E_{tot}/P_0)^{\frac{1}{3}}/C_0$ when C_0 is the pre-shock sonic velocity.

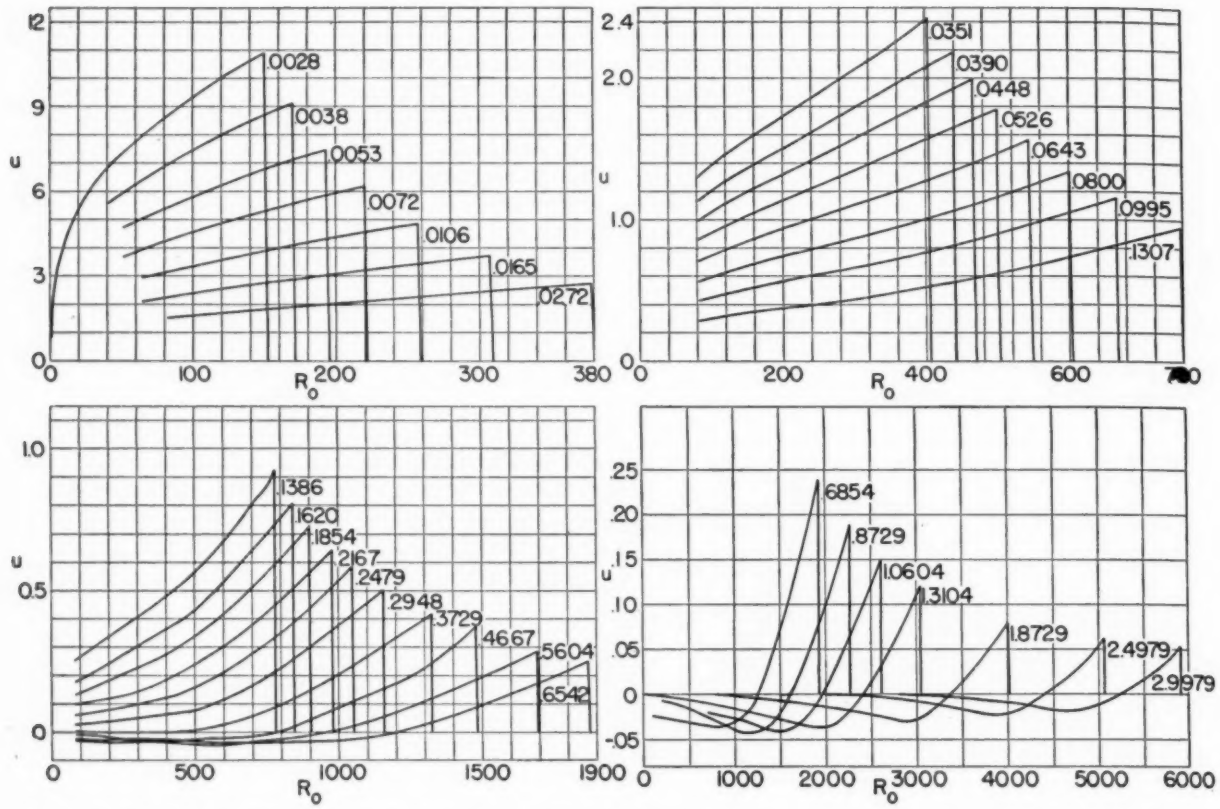


FIG. 5. Particle velocity in units of pre-shock sonic velocity as a function of Lagrange or mass position for the point-source solution at the times indicated. The position is in units of $(E_{tot}/P_0)^{\frac{1}{3}}/1627.2$, and the time is in units of $(E_{tot}/P_0)^{\frac{1}{3}}/C_0$.

The particle velocity follows its strong-shock dependence on shock radius down to as low an overpressure as $\frac{1}{10}$ atmos (Fig. 2); in fact an adequate fit over the entire range is given by the following:

$$u_s = 0.30\lambda_s^{-\frac{1}{2}}. \tag{19}$$

This simple form is reflected in the peak dynamic pressure ($Q_s = \frac{1}{2}\rho_s u_s^2$) which falls somewhat faster than an inverse cube law down to equally low shock pressures (Fig. 3). The slow decrease in shock density (from 6 for strong shocks to one for weak shocks) is responsible for the decay being steeper than the inverse cube of the shock radius.

Pressure (P) *versus* Lagrangean Distance (R_0)

The pressure behind the shock wave, shown in Fig. 4 as a function of the Lagrangean (or initial position) variable, retains the strong-shock form until quite low pressures. Note, for example, that the ratio of central pressure to shock pressure remains 37 percent down to 20 atmos and decreases slowly below that to 33 percent by 3 atmos. Beyond 3 atmos a negative phase develops with the pressure falling to as low as 0.8 atmos near the center.

In Fig. 4 the Lagrangean position is given in arbitrary units (R_0), related to the dimensionless unit (λ_0) by a constant multiplier,

$$r_0 = \epsilon\lambda_0 = \epsilon\frac{R_0}{1627.2}. \tag{20}$$

It may help in visualizing dimensions to assume R_0 to be in centimeters; then the corresponding blast energy will be about that of 200 lb of TNT (4.2×10^{15} ergs).

Particle Velocity (u) and Density (ϱ) *versus* Lagrangean Distance (R_0)

Figures 5 and 6 indicate the progression of particle velocity and density as functions of the Lagrangean position. Again the strong-shock form is dominant until as low as 3 atmos of shock overpressure.

The particle velocity transforms gradually from its very linear form to one much like the overpressure at large distances as the shock wave goes from strong to weak.

The density, which in the strong shock is zero at the origin, implying an infinite temperature, remains zero there since no mechanism is included (no conduction or radiation) for dissipating this high temperature.

The density profiles in Fig. 6 represent strong-shock initial conditions at 200 atmos shock pressure. The dip in the curves that sit at the same mass point ($R_0 = 150$) is due to the sudden inclusion of a finite

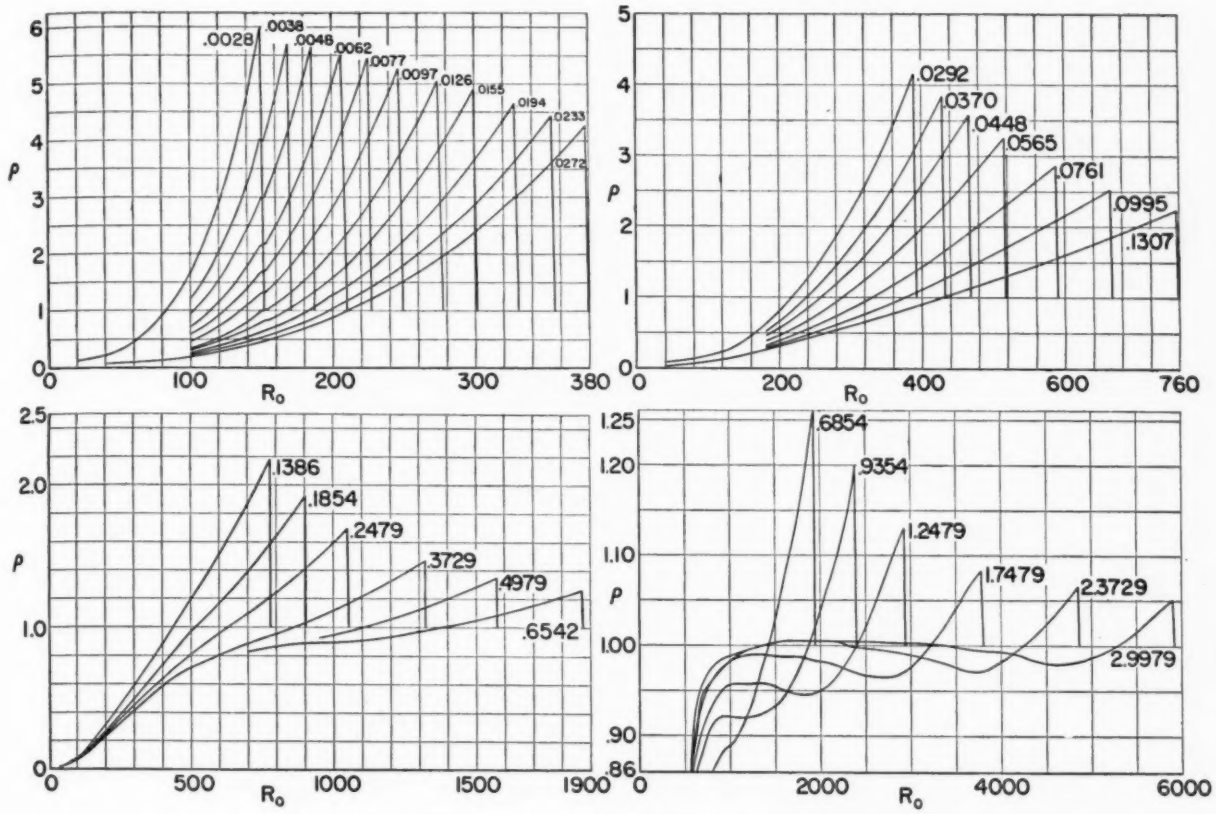


FIG. 6. Density in units of pre-shock density as a function of Lagrange or mass position for the point-source solution at the times indicated. The position is in units of $(E_{tot}/P_0)^{1/3}/C_0$.

atmosphere ahead of the shock, i.e., due to the abandonment of the strong-shock (infinite pressure ratio) assumption at that point. The Hugoniot relations give the shock density in terms of the shock pressure as

$$\rho=\frac{(\gamma+1)P+(\gamma-1)}{(\gamma-1)P+(\gamma+1)}=\frac{6P+1}{P+6} \quad \text{for} \quad \gamma=1.4. \quad (21)$$

For a strong shock ($P=\infty$) the density ratio is six ($\rho=6$), but at two hundred atmos ($P=200$) the ratio is only 5.83.

The strong-shock solution is derived on the assumption that the atmosphere ahead of the shock wave has negligible effect. The effect, when non-negligible, is to raise the temperature through the shock to a higher value than that given by the strong-shock solution, i.e., a finite shock is hotter than would be predicted by the "strong-shock" theory.

Gas Variables as a Function of Eulerian Position (λ)

The pressure, density, compression, and particle velocity are shown relative to their peak or shock values as functions of the Eulerian position (λ) in Fig. 7. The strong-shock form dominates the first two sets, while the later ones show the characteristic positive phase followed by a longer, weaker negative phase and eventually by a return to near pre-shock values at the origin.

Positive and Negative Phases *versus* Distance

The durations of the positive phases for pressure and particle velocity are shown in Fig. 8. Although these durations should approach the same value at large distances, they still differ by 7 percent at a distance of $\lambda=3.0(\Delta P_s=0.09)$. The figure indicates only the values for the point source solution.

The duration of the negative pressure phase (D_p^-), unlike that of the positive phase, is nearly independent of distance, and has an average value of 1.22. The negative durations suffer somewhat from loss of numerical significance at late times in the calculation.

Time Dependence of Pressure

In Fig. 9, the curves of pressure *versus* time at various distances are given in units of the peak overpressure and the positive duration in order to illustrate the change in the rate of decay behind the shock. At increasing distances the drop in pressure is both relatively and absolutely slower since the positive phase duration is also increasing. Again, these curves are for the point-source solution.

An exponential form approximates the time de-

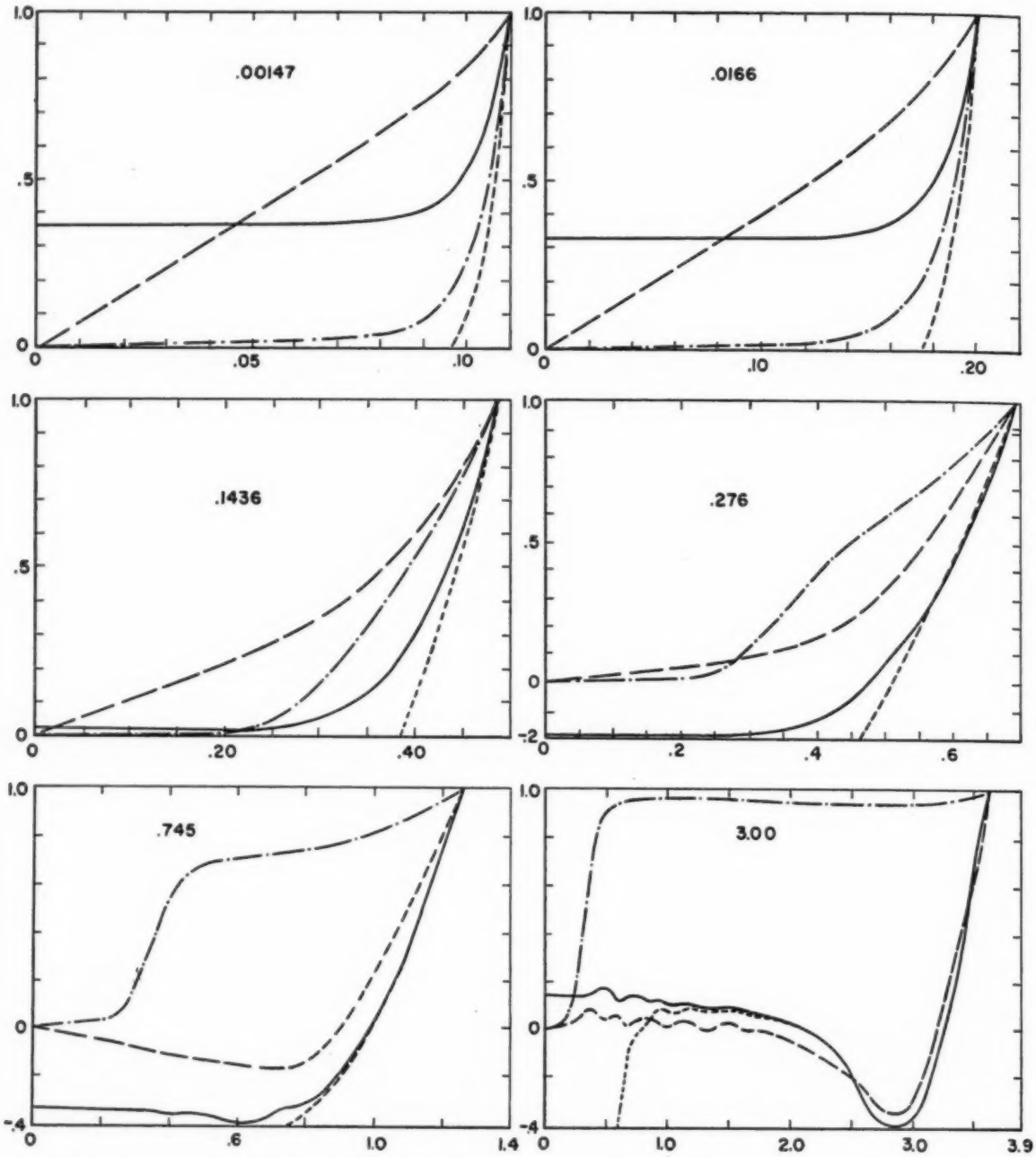


FIG. 7. Hydrodynamic quantities in units of their peak values as a function of Eulerian or space position at the times indicated. The radius is in units of $(E_{tot}/P_0)^{\frac{1}{3}}$. The solid curves represent overpressure $(\Delta P/\Delta P_s)$, the long dash curves represent particle velocity (u/u_s), the dot-dash curves represent density (ρ/ρ_s), and the short dash curves represent compression $[(\rho-1)/(\rho_s-1)]$. The shock values of the overpressure (ΔP_s) for these times are 121.5, 20.10, 2.03, 1.01, 0.338, 0.0701 ordered on increasing time. For the particle velocity the peak values are 8.49, 3.37, 0.873, 0.524, 0.216, 0.0525, and for the density the peak values are 5.66, 4.68, 2.135, 1.625, 1.228, 1.0508.

pendence of the pressure pulse:

$$\frac{\Delta P}{\Delta P_s}=(1-Z)\epsilon^{-\alpha Z}. \tag{22}$$

Here Z is the time after shock arrival in units of positive duration, ΔP_s is the shock overpressure, and α is independent of Z. This form is satisfactory for overpressures less than one atmosphere $(\lambda>0.74)$, with the coefficient (α) specified by

$$\alpha=\tfrac{1}{2}+\Delta P_s, \quad \Delta P_s\leqslant 1. \tag{23}$$

For shock overpressures greater than one atmosphere the decay is not a simple exponential, since the early portion requires a larger α than the later part. Allowing the coefficient α to be a function of the time (Z), it

may be approximated by

$$\alpha=\tfrac{1}{2}+\Delta P_s[1.1-(0.13+0.2\Delta P_s)Z] \qquad (24)$$

for ΔP_s less than 3 atmos, and for overpressures from 3 up to 50 atmos by the form

$$\alpha=a+\frac{b}{1+cZ} \qquad (25)$$

where

$$a=\begin{cases}-0.231+0.388\Delta P_s-0.0332\Delta P_s^2 & \text{for } \Delta P_s\leqslant 10\\ 0 & \text{for } \Delta P_s>10\end{cases}$$

$$b=\begin{cases}\Delta P_s(0.88+0.072\Delta P_s) & \text{for } \Delta P_s<10\\ \Delta P_s(1.67-0.011\Delta P_s) & \text{for } \Delta P_s>10\end{cases}$$

$$c=8.71+0.1843\Delta P_s-104/(\Delta P_s+10).$$

The time dependence of the pressure in the negative phase may be approximated by the form

$$\Delta P=14\Delta P_-\omega(1-\omega)\epsilon^{-4\omega} \quad \text{for} \quad 0.1<\Delta P_s<200 \qquad (26)$$

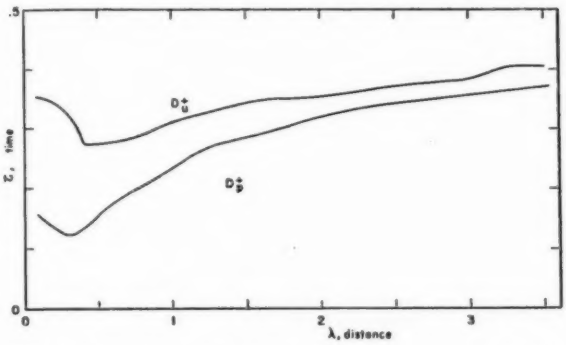


FIG. 8. Duration of positive phase for pressure (D_p^+) and particle velocity (D_u^+) *versus* distance (Eulerian) where time is in units of $(E_{tot}/P_0)^{\frac{1}{3}}/C_0$ and distance is in units of $(E_{tot}/P_0)^{\frac{1}{3}}$.

where ΔP_- is the peak negative overpressure (see Fig. 10), and where ω is the time measured from the end of the positive phase in units of the negative phase duration (D_p^-). A more accurate fit would allow the exponent to decrease to zero as the shock strength goes to zero.

Dynamic Pressure *versus* Time

Of interest also is the form of the *dynamic pressure* ($Q=\frac{1}{2}\rho u^2$). Time plots of this function appear in Fig. 11. These curves are also normalized, but with the positive duration of the particle velocity (D_u^+) (Fig. 8), and the peak value of the dynamic pressure (Q_s) (Fig. 3).

A similar exponential form approximates the dynamic pressure,

$$Q/Q_s=(1-Z)^2\epsilon^{-\beta Z}, \qquad (27)$$

where β is independent of the time (Z) for shocks of less than one atmos peak overpressure ($\Delta P_s<1$). The

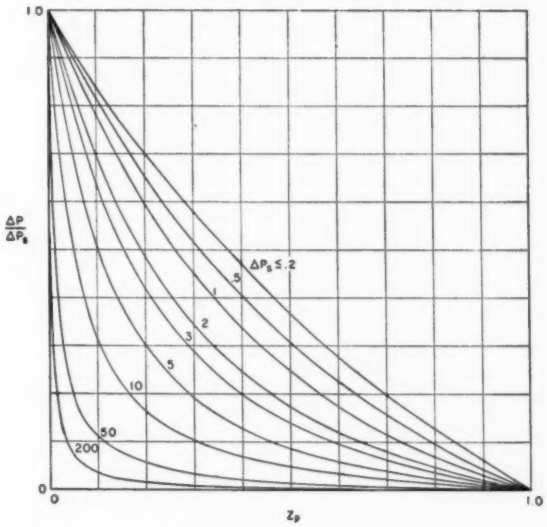


FIG. 9. Pressure as a function of time where the pressure is in units of the peak pressure (ΔP_s) (Fig. 1) and the time (Z_p) is in units of the positive duration (D_p^+) (Fig. 8). The numbers indicate the corresponding peak overpressures (ΔP_s).

coefficient β changes with the shock strength, however:

$$\beta=0.75+3.2\Delta P_s, \quad \Delta P_s\leqslant 1. \qquad (28)$$

Where the shock is stronger, a modification similar to that given in Eq. (25) for the overpressure is appropriate:

$$\beta=d+f/(1+gZ), \qquad (29)$$

for

$$50>\Delta P_s>1,$$

$$d=\begin{cases}-1.33\Delta P_s & \text{for } \Delta P_s\leqslant 3\\ -5.6+0.63\Delta P_s & \text{for } 3<\Delta P_s\leqslant 10\\ 0 & \text{for } \Delta P_s>10\end{cases}$$

and

$$f=6.40\Delta P_s,$$
$$g=0.725\Delta P_s.$$

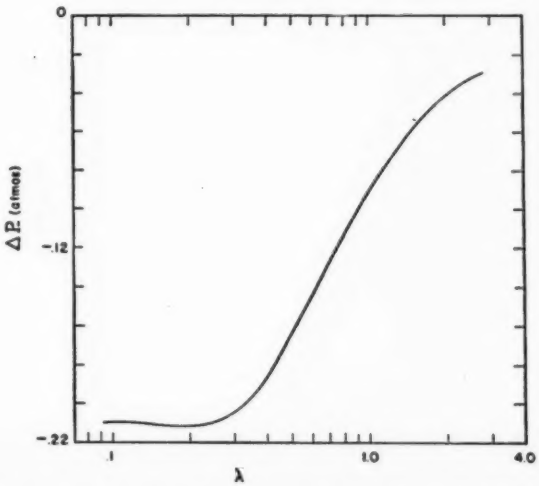


FIG. 10. Peak negative overpressure in atmos as a function of radial distance in units of $(E_{tot}/P_0)^{\frac{1}{3}}$.

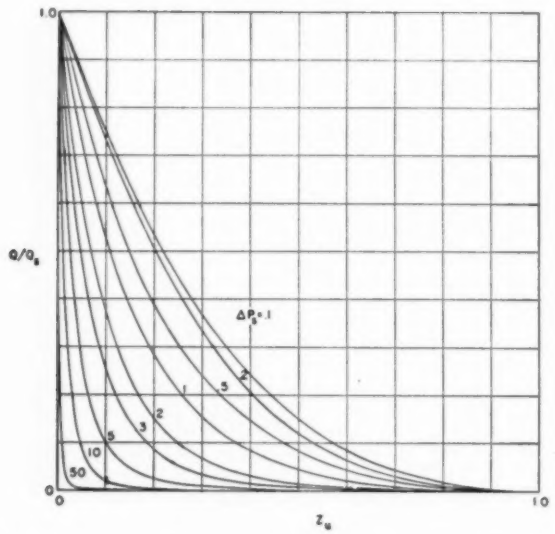


FIG. 11. Dynamic pressure ($Q=\frac{1}{2}\rho u^2$) as a function of time where the pressure is in units of the peak dynamic pressure (Q_s) (Fig. 3), and the time is in units of the positive duration of the velocity (D_u^+) (Fig. 8). The numbers indicate the corresponding peak overpressures (ΔP_s). This rapid decrease in effective pulse duration with increasing peak overpressure has the effect of making the higher dynamic impulses much the sharpest.

These approximate forms agree with the numerical values of Q and ΔP in the positive phase to within 10 percent (and over most of the range to less than 2 percent) for values of ΔP_s less than ten atmos.

Positive Impulse

The integrated positive overpressure (I_p^+) and the total positive drag pressure (I_u^+) decrease with distance from the blast source in the manner shown in Fig. 12.

$$I_p^+=\int_0^{D_{p^+}}\Delta P(t)dt,$$
$$I_u^+=\tfrac{1}{2}\int_0^{D_{u^+}}\rho u^2 dt. \tag{30}$$

(Because of the units employed in this paper, I_u^+ does not become the usual dynamic impulse until it is multiplied by $\gamma=\rho_0 C_0^2/P_0$.)

Both of these impulses may be fitted by simple powers of the radial distance for shock overpressures less than two atmospheres:

$$I_p^+=0.043\lambda^{-1},\quad \Delta P_s<2$$
$$I_u^+=0.004\lambda^{-2.7}. \tag{31}$$

Negative Impulse

The negative overpressure impulse in the range below 20 atmos of peak overpressure can be expressed to within 5 percent by

$$I_p^-=\tfrac{1}{2}\Delta P_- D_p^-. \tag{32}$$

Since the duration of the negative pressure is constant at $D_P^-\simeq 1.22$,

$$I_P^-\simeq 0.61\Delta P_-. \tag{33}$$

For distances greater that $\lambda=0.5$ the negative peak overpressure approaches zero inversely as the distance

$$\Delta P_-\simeq -0.086\lambda^{-1},\quad \lambda>0.5, \tag{34}$$

and the negative pressure impulse goes approximately like

$$I_P^-\simeq 0.052\lambda^{-1},\quad \lambda>0.5. \tag{35}$$

Special Features of the Isothermal Sphere Problems

In the gas dynamics resulting from the release of an initially static high-pressure sphere, an inward moving shock forms behind the rarefaction wave that first runs in from the surface of the sphere. This inward-directed shock was predicted by Wecken[7] and discussed by McFadden,[8] Shardin[9] and others.[4] It does not acquire a net inward velocity until the rarefaction has reached the center, but after that it moves in and reflects at the origin, and then races outward to eventually overtake the main shock.

This second shock grows from zero strength to presumably an infinite pressure ratio at the origin. On reflection, this shock moves outward, decaying in strength about as the inverse first power of its distance

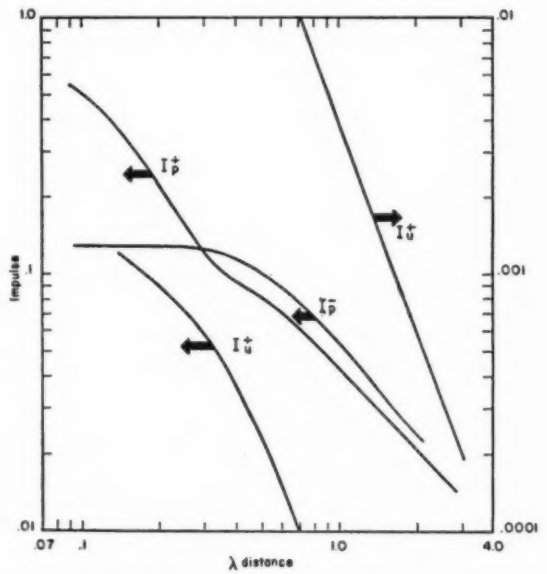


FIG. 12. The radial dependence of the positive pressure impulse (I_p^+) the negative pressure impulse (I_p^-), and the positive dynamic impulse I_u^+ in units of atmospheres [where time is in the dimensionless units $(E_{tot}/P_0)^{\frac{1}{3}}/C_0$] the radial distance is in units of $(E_{tot}/P_0)^{\frac{1}{3}}$.

[7] F. Wecken, Z. angew. Math. u. Mech. **30**, 270 (1950).
[8] J. A. McFadden, J. Appl. Phys. **23**, 1269 (1952).
[9] H. Schardin, "Measurement of spherical shock waves," Commun. Pure Appl. Math. (Institute of Mathematical Sciences, New York University) **VII**, 223 (1954).

from the origin. The general expansion and outward motion of the gas behind the main shock is responsible for this apparently modest decay rate.

When this second shock encounters the outer surface of the gas that was initially inside the sphere, a transmitted shock continues out and a reflected shock is sent inward. The transmitted shock overtakes the main shock and increases it by as much as 20 percent. Progressively weaker shocks follow from the reflections at the interface and again at the origin. This repeated shocking of the gas near the origin changes the temperature profile at later times from its initial isothermal nature to nearer the point-source distribution which is characterized by a high central temperature falling off rapidly with radius.

Some detail is lost in the vicinity of the origin since the shocks are always spread over a number of mesh points. In one case the problem was rerun with mesh sizes about one-fourth the original size, and the inward-moving shock showed some appreciable discrepancy near the origin. However, this discrepancy may be attributed mainly to the difficulty in identifying the shock when the rounding is comparable to the shock radius. After reflection the difference vanishes again in spite of the quite different histories near the center.

IV. CONCLUSIONS

Point-Source Solution

At increasing distances from a finite but sudden source of energy, the resulting blast wave will appear more and more like that from a point source. The blast resulting from an initial isothermal sphere of gas at rest will assume the general shape and values of the point-source solution (to within 10 percent) after the shock wave has engulfed a mass of air 10 times the initial mass of the sphere. Prior to this, the shock strength is less than that of the point-source shock, and the inward-traveling rarefaction has not reached the center.

A point source should leave a higher temperature and consequently a larger percentage of energy near the origin. This energy, no longer available to the shock wave, should effect a reduction in the shock radius (for a given overpressure). However, no appreciable difference in the low-end of the shock overpressure-radius relation appears on comparing the point source and isothermal sphere solutions. In fact, perhaps because of multiple shocking of the inner regions in the isothermal sphere problems, the distribution of residual energies (per unit volume) and pressures are nearly identical around the origin at a time when the shock has progressed to 6 or 7 times the initial radius. Although temperatures will remain different, since the point source has infinite temperature at the origin (and zero density), the average temperature (or density) of that gas initially inside the isothermal sphere approaches (within 10 percent) the average temperature for a corresponding mass around the point source.

For a source of considerable initial mass, the peak pressure may become quite small before the shock has engulfed a mass of gas 10 times larger, and the blast wave may, therefore, remain quite different from the point-source solution throughout regions of interest. Such is the case to some extent with high explosives where initial charge shapes will influence the blast wave at all significant pressures. It is true to an even greater extent in the spherical equivalent of a shock-tube type of blast, where a gas at high pressure but at normal temperature is suddenly released, as in the normal temperature isobaric sphere problem (dotted curve Fig. 1) described in Sec. I.

Application to Blast in Air

The ideal gas assumption is reasonably valid in air for shock pressures less than 10 atmos. Above that the gamma in the expression for the internal energy [Eq. (7)] ranges down to as low as 1.133 and up to the monatomic value of 1.667. But since the mass of gas which has experienced shocks stronger than 10 atmos is a small part (5 percent) of the engulfed mass by the time the shock overpressure is down to 1 atmos, the blast wave for constant gamma should be reasonably correct for air at most interesting shock pressures.

In the case of shock wave problems involving only one space variable, the artificial viscosity technique for numerical integration appears to be very satisfactory over large ranges of pressure and entropy change. Unfortunately, it cannot be expected to yield details of a shock wave impinging on a singular point such as the origin in spherical or cylindrical geometry, since the nature of this method is such as to spread the shock front over a number of grid points.

ACKNOWLEDGMENTS

The persons participating in and aiding the work reported here are too numerous to list; however, at the risk of slighting many, I mention a few: R. Latter gave some critical attention to the formulation of the problem. Ruth Anne Engvall, I. Greenwald, and other mathematicians performed much of the coding and numerical work. Miss Engvall also supervised much of the analysis of results and the preparation of curves. In addition, the author enjoyed frequent helpful discussions with other RAND personnel and consultants.

This work was accomplished with the aid of both Air Force and U. S. Atomic Energy Commission funds.

Letters to the Editor

Method for Producing High-Velocity Metallic and Plastic Pellets*

M. E. VAN VALKENBURG AND CHARLES D. HENDRICKS, JR.
University of Utah, Salt Lake City, Utah
(Received November 15, 1954)

THE acceleration of metallic pellets to velocities of 2–6 km/sec has been reported by Allen, Rinehart, and White[1] using aluminum, magnesium-lithium, titanium, and steel for pellet materials. In their work, the pellet was embedded in the end of a cylindrical explosive charge. Detonation of the charge caused the pellet to be propelled in the direction along the axis of the cylinder. However, the pellet frequently broke into a number of pieces at the time of firing or was completely shattered in the case of some metallic pellets.

Kolsky, Snow, and Shearman[2] have reported the use of water, powdered chalk, and other inert materials to dampen the shock-wave amplitude in their study of the deformation of shaped charge liners. By the use of inert materials such as oil or plastics as a coupling medium between the pellet and the explosive, the method of Allen, Rinehart, and White has been modified for use with other metals, other pellet shapes, and for a variety of plastic materials.

Details of the construction of the propelling mechanism are shown in Fig. 1. The pellet is suspended a distance from the explosive charge on oil paper. The space between the oil paper and explosive charge is filled with oil or other material. The acceleration of the pellet is the result of both the shock wave and the mass motion of the explosive products following detonation. Photographs taken with a high-speed framing camera (2 μsec per frame) show that the pellet lags behind the shock wave and is completely surrounded by explosive products before it moves appreciably from its initial position. The pellet velocity depends on the mass and cross-sectional area of the pellet, the type of explosive charge, and the material and thickness of the coupling medium.

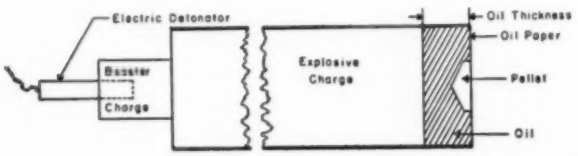


FIG. 1. Assembly of pellet and oil or plastic coupling to an explosive charge for accelerating pellets. Typically, the explosive charge is 2½ in. in diameter and 7¼ in. in length for a 1 in. diameter pellet.

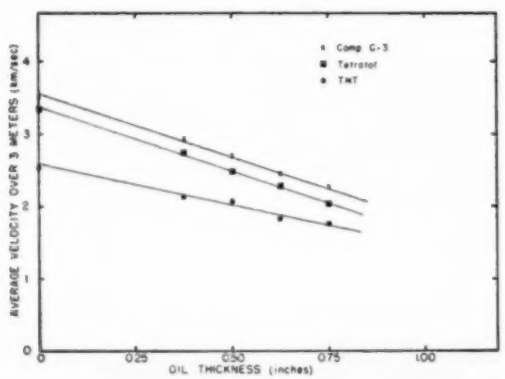


FIG. 2. Average pellet velocity as related to oil thickness for aluminum pellets (5.25 grams) using castor oil as the coupling medium.

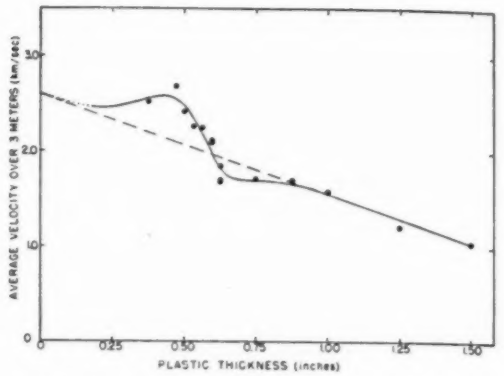


FIG. 3. Average velocity of aluminum pellets using a cast polyester resin as a coupling medium. The straight dashed line indicates the velocity of pellets with castor oil as the coupling medium as given in Fig. 2.

Breakup of pellets has been tested on thin targets made of aluminum foil and paper. In the case of no breakup, there is one hole which has regular boundaries and is of a characteristic shape (some projection of pellet shape). With breakup, however, there are a number of irregular shaped holes for each shot. By the use of the interposed layer of inert material, both pellet breakup and pellet velocity can be controlled by the thickness and material in the layer.

In Fig. 2, the average pellet velocity over the first three meters of pellet flight is shown as a function of oil thickness. These measurements were taken with castor oil as the coupling medium. For this case, there is an essentially linear decrease in velocity with oil thickness.

The use of a cast plastic layer between the pellet and explosive shows somewhat different results. Figure 3 shows average pellet velocity as a function of plastic thickness. The plastic material used was a polyester resin known commercially as "Aqua Armor." In this case, a resonance phenomena occurs with a thickness of about 0.5 inch. The velocity of the pellet is higher than with an oil interposed layer using the same pellet materials and the same explosive charge.

Figure 4 shows the average velocity-layer thickness characteristics of several plastic materials. Because of their smaller mass for the same area, plastic pellets attain higher velocities than metallic pellets. Pellets that have been accelerated without breakup include phenolic-bonded fabrics, potted epoxy resins (such as Epon 828), Lucite, polystyrene, Textolite, nylon embedded in polyamide, and Teflon.

We wish to acknowledge use of the test facilities of the Ex-

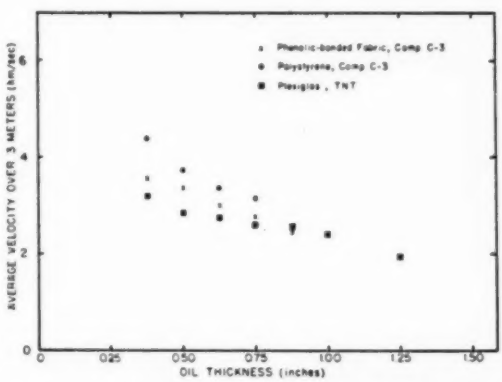


FIG. 4. Average velocity of plastic pellets as a function of oil thickness for phenolic-bonded fabric pellets (2.75 grams), polystyrene pellets (2.34 grams), and Plexiglas pellets (2.53 grams). In each case, castor oil was the coupling medium.

plosives Research Group, University of Utah, under M. A. Cook, and the assistance of Wallace G. Clay and Philip Weinberg in conducting the experiments.

* This research was supported by the Office of Ordnance Research, U. S. Army.

[1] Allen, Rinehart, and White, J. Appl. Phys. **23**, 132 (1952).
[2] Kolsky, Snow, and Shearman, Research **2**, 89 (1949).

Collapsed Toroidal Springs as Contact Elements in Microwave Cavities

C. L. Andrews
General Electric Research Laboratory, Schenectady, New York
(Received November 26, 1954)

CYLINDRICAL spiral springs welded together at the ends to form toroids have been studied as contact elements in coaxial microwave cavities. The first ones tried were made from stainless steel wrist bands of the five and ten cent store variety. Stainless steel was spot-welded together at the ends easily without being made brittle. The springs of flattened oval shape made long line contacts with cylindrical cavities.

However, the sizes of stainless-steel wrist bands were limited. It was not possible by any simple means to wind stainless steel on mandrels of oval cross section and remove the spring from the mandrel. If a spring was wound on an oval mandrel without heat treatment, it would partially unwind upon removal so that the major axes of the ovals were not parallel.

A surprising solution to the problem came when it was found that springs wound in cylinders of circular cross section could be collapsed to give springs of oval cross section like that of the wrist band.

Figure 1 shows how the toroid may be collapsed by rotating the outer circumference relative to the inner circumference of the toroid. The two sets of concentric cylinders are shown to indicate the wide tolerance permitted in diameters of the cylindrical parts. Identical springs are used in the two cases. The illustration is an exaggeration for emphasis. The more the spring is collapsed the longer will be the line of contact between the wire and the cylindrical wall.

Figure 2 illustrates the wide range in eccentricity of cylinders that may be tolerated. It exaggerates the eccentricity often found in disk-seal triodes.

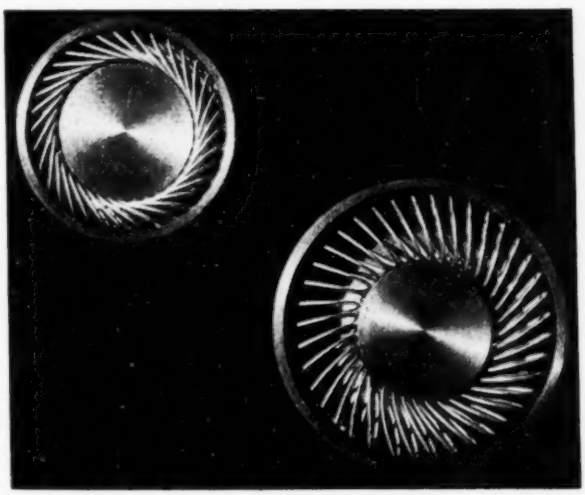


FIG. 1. Collapsed toroidal contactors indicating the tolerance permitted in cylindrical parts.

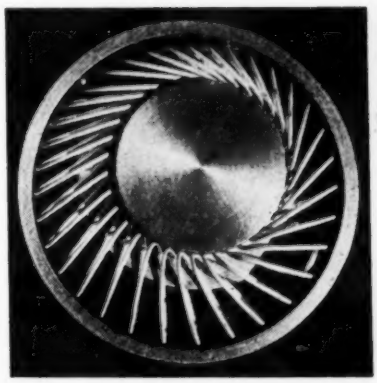


FIG. 2. Illustration of tolerance in eccentricity.

Although the toroid springs were easily collapsed with the fingers to fit into a smaller cylinder, they were somewhat more convenient if they are permanently collapsed so that the thickness of the oval was only 20 percent greater than when in use. The collapsed springs were formed by heating the toroids to the surprisingly low temperature of 220°C in a hydrogen atmosphere between two cylinders for only 20 sec. The expression "feathered toroid" may be more descriptive than "collapsed toroid" since the individual turns lie over each other like the feathers of a bird.

When a toroid was mounted on an inner cylinder and collapsed by pressing radially inward with the fingers, the radial force was observed to be greatest when the turns lay in nearly radial planes. The force decreased to a minimum as the spring collapsed and then increased as the individual turns began to touch each other near the inner cylinder.

The radially outward force by one turn was measured by a compression balance inserted through holes in the sides of cylinders, such as those of Fig. 1. Six outer cylinders were used ranging

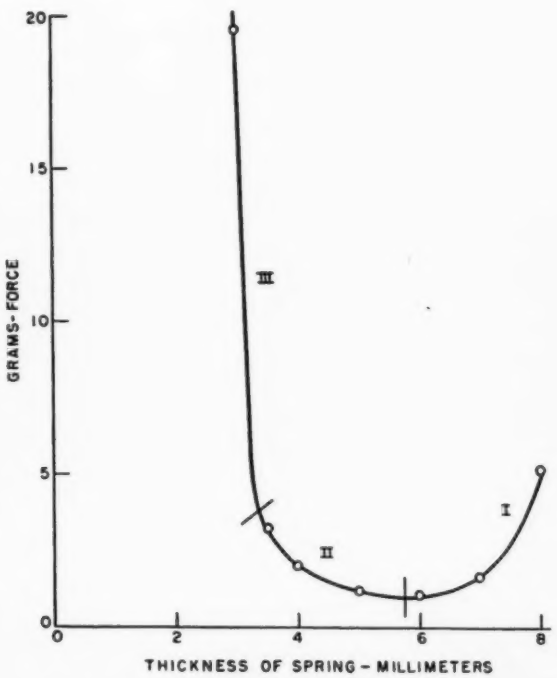


FIG. 3. Radial force *versus* thickness of the collapsed spring.

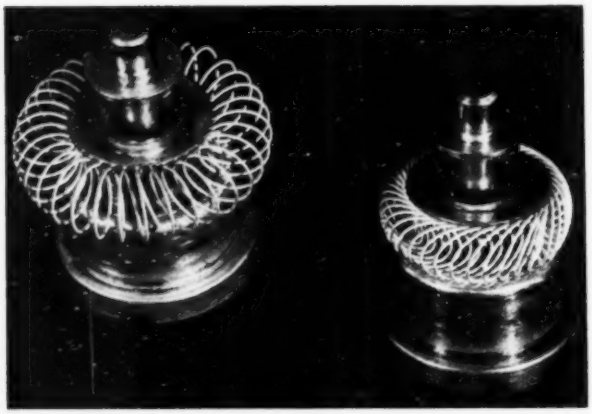


FIG. 4. Simple toroid and collapsed toroid mounted on disk-seal triode tubes.

in diameter from $1\frac{1}{16}$ to $1\frac{3}{8}$ in. in diameter. The individual turn was pushed in with the compression balance until the outer edge of that turn lay in the same circle as the outer edge of the other turns. A Starrett spring gauge of 0.001-in. scale divisions was calibrated to serve as a compression balance.

The toroidal springs on which these measurements were made were of 0.002-in. wire, 40 turns in a cylinder of 0.37 outside diameter, spot-welded together at the ends and mounted over a cylinder of $\frac{1}{16}$ in. diameter. The separation of turns along the inner cylinder equals roughly the wire diameter.

Figure 3 is a graph of radial force in grams against thickness of the spring; that is, the difference in radii of inner and outer cylinder. The useful cases are indicated by II and III on the graph. Over range II each turn of wire is a circular loop and the cross section cut by a plane including the axis of the cylinder is an ellipse. In part III each turn of wire is bent over the next one. The ellipse is changed to a flattened oval and the wire makes a longer line of contact with the cylinder. The end view of the toroid appears as almost continuous metal.

Figure 4 shows a simple toroid and a collapsed toroid mounted for use on 2C40 disk-seal (lighthouse) triodes. In this position, the spring makes three sets of contacts: butt contacts to the cathode cylinder, point contacts to the grid, and line contacts to the wall of the cylindrical cavity into which the tube will be inserted. The spring completely closes off the grid-cathode cavity. This serves in oscillators of wavelength 12–13 cm. The feedback is entirely within the tube. In previous oscillators in which the connectors to grid and cathode were made by milling slots in the brass cylinders, the output of the oscillator had varied widely as the tube was rotated about its axis. This was due to eccentricity in tubes. When the collapsed toroidal spring contactors were used, no change in output of the oscillator was observed as the tube was rotated.

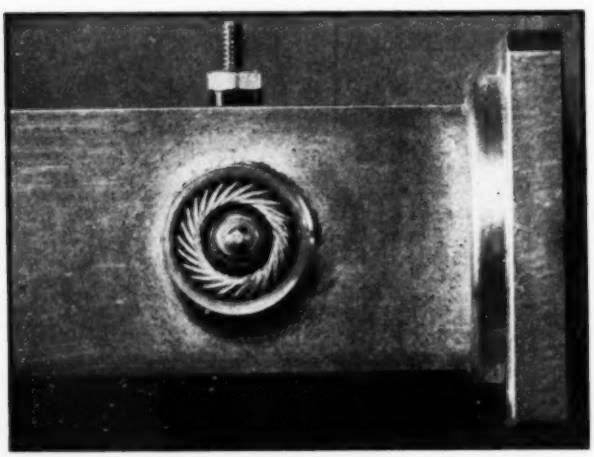


FIG. 5. Collapsed toroidal spring used as cathode contactor in rectangular wave-guide oscillator.

The Q's of plate-grid cavities with brass fingers ranged from 390 to 430 at 2400 Mc/sec. The Q's with the same tubes and sizes of cavity employing collapsed toroidal spring contactors were from 300 to 360.

The Q's were also measured in coaxial cavities without tubes to see if the springs would be suitable in measuring equipment. The ones of silver plated coaxial cavities $\frac{1}{2}$ in. inside diameter and 1 in. outside diameter, and a half-wavelength long, and shorted at both ends were from 1400 to 1500 at 2400 Mc/sec. When one end of the half-wave cavity was terminated with a collapsed toroid the Q was from 1100 to 1200.

Figure 5 shows a collapsed toroidal spring used as cathode contactor in a rectangular wave-guide oscillator employing a GL-6299 triode. This oscillator is used to generate microwaves of wavelength 5–6 cm.

The springs described thus far were made with stainless steel. Beryllium copper springs made equally good contacts. They were spot-welded with a silver lead eutectic solder at 304°C. The beryllium copper toroids could not be permanently collapsed by heat treatment. However, they were easily collapsed in the fingers before being inserted into the cavity. Upon removal they returned to the original shape. The beryllium copper was silver plated directly. The stainless-steel springs were plated with copper followed by silver.

The collapsed toroidal springs have the same cross section as wrist band contactors. Their tolerances are greater than that of wrist bands or contactors of machined fingers.

Further Remarks on the Odd Ball Problem as an Example in Information Theory

F. H. MITCHELL AND ROBERT N. WHITEHURST
University of Alabama, University, Alabama
(Received January 6, 1955)

IN a recent Letter to the Editor,[1] an interesting analysis of the odd ball problem is presented from the standpoint of information theory. The problem is to sort out from a number of balls, usually 12, the one which is different in weight, and to determine whether it is heavy or light, using an equal-arm balance. For the 12 balls, the number of weighings required is three.

The information required to solve the problem is shown to be $\ln 2N$, where N is the number of balls. The most information that can be counted on per weighing is $\ln 3$. Thus, as stated in the letter referred to, the number of weighings required is not fewer than $n=\ln 2N/\ln 3$, from which $N=3^n/2$. Since N is restricted to integral values, it is limited to $(3^n-1)/2$, which is the next smaller integer.

It is interesting to inquire as to the condition under which the problem is solvable when $N=(3^n-1)/2$. This number is of the form $3k+1$, where k is an integer. For the first weighing, the best choice is to place k balls in each balance pan, leaving out $k+1$ balls. The most unfavorable result in terms of information gain is for the pans to balance, yielding information equal to $\ln 3N/(N+2)$. In the remaining $n-1$ weighings, it is not possible to count on more information than $\ln 3^{n-1}$. Thus, the total information gain in the n weighings becomes $\ln 3^n N/(N+2)$, which is equal to $\ln 2N(N+\frac{1}{2})/(N+2)$. Since this quantity is less than $\ln 2N$, the problem is not solvable when $N=(3^n-1)/2$.

Suppose, however, that in addition to the N balls, there is available a single good ball. The problem is now solvable. For the first weighing, place $k+1$ balls in one pan, and k plus the good ball in the other pan, leaving k balls out. It is not difficult, though tedious, to show that the information gain is now sufficient to solve the problem with $N=(3^n-1)/2$.

It follows immediately that $3N$ balls can be solved in $n+1$ weighings. Since $3N$ is divisible by 3, the first weighing can furnish information equal to ln3, and will also furnish the good ball necessary for the next weighing. Therefore, $(3^n-3)/2$ balls can be solved in n weighings.

These same results, with proof based on mathematical induction rather than information theory, can be found in the mathematical literature.[2,3]

[1] Paul J. Kellogg and Dorothy J. Kellogg, J. Appl. Phys. **45**, 1438 (1953).
[2] N. J. Fine, Am. Math. Monthly **54**, 489 (1947).
[3] Math. Gazette **30**, 231 (1946).

Potential Distribution as a Function of Current in the Spherical Diode

ANNIJA DUNKULS AND EARL ZWICKER
Department of Physics, Illinois Institute of Technology, Chicago, Illinois
(Received January 19, 1955)

USING the method developed by Copeland and Eggenberger[1] the functions used in the calculation of potential distribution in the case of the spherical diode have been evaluated by direct integration and tabulated. The equations resulting from this analysis are

$$\left(\frac{dV}{dr}\right)_c=\left(\frac{dV}{dr}\right)_{co}(1-a\delta-b\delta^2-\cdots)$$

and

$$\left(\frac{dV}{dr}\right)_a=\left(\frac{dV}{dr}\right)_{ao}(1+a'+b'^2+\cdots),$$

where

$$\delta\equiv I/I_s,$$

(I representing the actual current and I_s the space-charge-limited current). Derivatives as shown are evaluated at the cathode if the subscript is c or at the anode if the subscript is a. Quantities evaluated in the absence of space-charge are indicated by the additional subscript o. The coefficients of the current ratio δ vary with the ratio of the cathode radius to the anode radius as shown in Table I. The range of values covered by the table includes both the external anode and the external cathode. The last column gives the ratio of the off-anode field for space-charge-limited current to the off-anode field in the absence of current.

TABLE I.

$x=r_c/r$	a	b	a'	b'	$\frac{[(dV/dr)_a]_{\delta=1}}{(dV/dr)_{ao}}$
500	0.5967	0.1106	0.0011	0.00007	1.0010
200	0.5963	0.1104	0.0027	0.0002	1.0024
100	0.5958	0.1103	0.0053	0.0003	1.0050
50	0.5949	0.1100	0.0101	0.0007	1.0098
20	0.5937	0.1097	0.0234	0.0016	1.0259
10	0.5935	0.1097	0.0434	0.0030	1.0483
5	0.5930	0.1097	0.0788	0.0056	1.0879
1	0.59259	0.10974	0.29630	0.02195	1.3333
0.50	0.5928	0.1097	0.5136	0.0377	1.5771
0.20	0.5937	0.1097	1.0551	0.0747	2.1657
0.100	0.5947	0.1097	1.8195	0.1229	3.0203
0.050	0.5962	0.1096	3.1543	0.2007	4.4743
0.020	0.5987	0.1095	6.607	0.3821	8.196
0.0100	0.6008	0.1094	11.686	0.624	13.616
0.0050	0.6032	0.1094	20.875	1.027	23.362
0.0020	0.6062	0.1094	45.58	2.011	49.366
0.0010	0.6083	0.1093	83.05	3.380	88.592

The results obtained confirm the conclusions of Copeland and Eggenberger[1] in the cylindrical case. They show that the ratio of the off-cathode field to the field in the absence of space-charge is not a universal function of the ratio of the actual current to the space-charge-limited current, but that it is a function which varies slowly as the geometry is changed.

[1] P. L. Copeland and D. N. Eggenberger, J. Appl. Phys. **43**, 280 (1952).

Measurement of Wall Temperature and Heat Flow in the Shock Tube*

ALBERT J. CHABAI AND RAYMOND J. EMRICH
Lehigh University, Bethlehem, Pennsylvania
(Received January 31, 1955)

MEASURED values of flow parameters (such as shock speed, density, and flow Mach number) deviate from those predicted by the ideal shock tube theory, which neglects friction and heat transfer. These deviations suggest that there is considerable heat transfer between the flowing gas and the walls of the shock tube. This letter reports a device to measure the heat flow, and presents some preliminary results.

The device, a gold foil which acts as a resistance thermometer, measures the temperature at the surface of the shock tube wall. Because heat penetrates only a small distance during the approximately 30 milliseconds of primary shock tube flow, the wall can be treated as a semi-infinite medium. The heat flow through the surface can be calculated from the measured variation of surface temperature with time. The device is calibrated for heat flow measurement by observing the temperature rise of the surface when a known heat input is supplied by establishing a current in the gold foil; thus, it is not necessary to assume values of thermal conductivity and diffusivity of the medium.

The gold foil is so thin (approximately 0.000003 inch) that it assumes the temperature of the surface on which it is mounted in considerably less than 1 μsec. The foil is mounted with shellac on 0.0015 inch-thick linen paper[1] and imbedded with a thermosetting resin[2] in an 0.008 inch-deep groove milled into the duralumin shock tube wall. The 0.008 inch-thick matrix of paper and resin is the semi-infinite medium referred to. Leads to the thermometer foil are heavier gold foils pressed into contact with the thermometer foil upon assembly of the shock tube section. The thin foil has a resistance of about 10 ohms.

Changes in resistance of the foil are observed by oscillographic recording of the changes in electrical potential across the foil through which a constant current is maintained. The temperature coefficient of resistance is determined by bridge measurement of resistance when the shock tube section containing the foil is placed in an oven.

There is a slight change in resistance of the foil associated with an isothermal pressure change in the shock tube. The resistance change produced by a 4-atmos isothermal increase in pressure is comparable to that associated with an isobaric 0.1°C temperature change. In the preliminary studies, the resistance changes observed are interpreted as caused solely by temperature changes, since the pressure changes were less than one atmos in all cases, and the observed resistance changes were many times the amount expected owing to pressure change.

Transient wall temperatures have been studied with nitrogen in a metal-walled shock tube of $\frac{3}{8}$ in. by 4 in. cross section at various distances from the diaphragm in both chamber and channel. These studies are preliminary but show that a variety of phenomena can be measured. Typical of these preliminary results are those obtained at 3.6 m from the diaphragm. Observations at this distance in the channel are summarized in Table I. The heat flow and temperature values are reliable only as an order of magnitude. The heat flow values there are representative of measurements during the first 30 μsec after the shock passes over the foil; the heat flow decreases somewhat at later times, but is still in the

TABLE I. Wall temperature rise and heat flow in shock tube channel.

Initial channel pressure	¼ atmos	¼ atmos	1/10 atmos
Initial chamber pressure	1 atmos	5 atmos	5 atmos
Heat flow to wall	1.5 w/cm²	6 w/cm²	7 w/cm²
Maximum temperature rise of wall	>0.6°C	8°C	7°C
Time after shock arrival that wall temperature is maximum	>10 millisecond[a]	7 millisecond	9 millisecond
Predicted time after shock arrival that cold front arrives	32 millisecond	6 millisecond	2.5 millisecond

[a] Observation of primary flow is terminated at this time by arrival of reflected wave from closed end of tube.

direction from gas-to-wall even after the wall temperature reaches the maximum values indicated. This continued flow of heat to the wall persists for a considerable time after the cold front is predicted to arrive on the basis of the ideal shock tube theory.[3] This indication of "frictional heating" of the gas in the boundary layer is confirmed by studies in the shock tube chamber. The wall temperature *increases* in the flow created by the rarefaction even though the gas temperature (and the stagnation temperature) is, according to the ideal theory,[3] below the wall temperature. With initial chamber pressure 5 atmos, and initial pressure ratios of both 10 and 50, the wall temperature gradually rises 2 or 3 degrees during the first 3 millisecond of flow. It then decreases monotonically to several degrees below the starting value during the next 10 millisecond. The primary flow is interrupted at that time by a reflection from a closed end of the shock tube.

Improvements in the construction of the heat flow gauge and in the calibrating and recording circuitary can be expected to yield reliable and detailed information on heat transfer to the walls in transient gas flows.

* Supported by the Office of Naval Research.
[1] The gold foil mounted on paper was kindly supplied by W. F. Grupe of the Peerless Roll Leaf Company, Union City, New Jersey.
[2] Scotch Weld No. 588 manufactured by Minnesota Mining and Manufacturing Company.
[3] For a description of the ideal theory, see, e.g., Glass, Martin, and Patterson, UTIA Report No. 2, University of Toronto Institute of Aerophysics (1953).

Generalizations of Brillouin Flow

L. R. WALKER
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received February 10, 1955)

IT is possible to generalize in a simple way some of the familiar self-consistent Brillouin flows of electrons in a uniform dc magnetic field.[1] Suppose this field to be of magnitude, H_0, and to be directed along the z-axis. Suppose that by some means the flow is established in such a way that the constant sum of potential and kinetic energies is the same for each electron and that the value of this sum is zero. Any motion in the z-direction will be considered as uniform and may be ignored. Then

$$\dot{x}^2+\dot{y}^2+\frac{2e}{m}V=0,$$

where $\dot{x}$ and $\dot{y}$ are the x and y velocities and V is the electrostatic potential. In a steady flow of electrons, we may replace d/dt in the equations of motion by $\dot{x}(\partial/\partial x)+\dot{y}(\partial/\partial y)$; doing this one finds

$$\frac{\partial\dot{x}}{\partial y}-\frac{\partial\dot{y}}{\partial x}=\omega_H$$

where $\omega_H=e\mu_0H_0/m$ and μ_0 is the permeability of free space. The remaining equations are that of continuity and Poisson's equation.

It is convenient to introduce in place of x and y the variables, $\sigma=x+jy$ and $\sigma^*=x-jy$. Then the available equations are easily verified to be

$$V=-\frac{m}{2e}\dot{\sigma}\dot{\sigma}^*,$$

$$\frac{\partial^2V}{\partial\sigma\partial\sigma^*}=\frac{\rho}{4\epsilon_0},$$

$$\frac{\partial\dot{\sigma}}{\partial\sigma}-\frac{\partial\dot{\sigma}^*}{\partial\sigma^*}=-j\omega_H,$$

$$\frac{\partial\rho\dot{\sigma}}{\partial\sigma}+\frac{\partial\rho\dot{\sigma}^*}{\partial\sigma^*}=0,$$

where ρ is the charge density. Suppose one now asks that ρ be constant. Then from the last two equations,

$$\frac{\partial\dot{\sigma}}{\partial\sigma}=-\frac{j}{2}\omega_H$$

and, hence

$$\dot{\sigma}=-\frac{j}{2}\omega_H\sigma+f(\sigma^*),$$

where f is an arbitrary function. Similarly,

$$\dot{\sigma}^*=\frac{j}{2}\omega_H\sigma^*+f^*(\sigma)$$

and

$$V=-\frac{m}{2e}\left[\frac{j}{2}\omega_H\sigma^*+f^*(\sigma)\right]\left[-\frac{j}{2}\omega_H\sigma+f(\sigma^*)\right].$$

Substituting in Poisson's equation, one finds

$$f'(\sigma^*)f^{*\prime}(\sigma)=\frac{\omega_p^2}{2}-\frac{\omega_H^2}{4},$$

where $\omega_p^2=e\rho/m\epsilon_0$. The only solutions of this equation are

$$f'(\sigma^*)=\left(\frac{\omega_p^2}{2}-\frac{\omega_H^2}{4}\right)^{\frac{1}{2}}e^{j\alpha},$$

where α is real and $\omega_p^2>\omega_H^2/2$. Thus, the possible velocities are given by

$$\dot{\sigma}=-\frac{j}{2}\omega_H\sigma+\left(\frac{\omega_p^2}{2}-\frac{\omega_H^2}{4}\right)^{\frac{1}{2}}e^{j\alpha}\sigma^*+\text{constant}.$$

By a suitable translation and rotation of axes this may be brought to the form (retaining σ and σ^* as labels for the new variables)

$$\dot{\sigma}=j\left[-\frac{\omega_H}{2}\sigma+\left(\frac{\omega_p^2}{2}-\frac{\omega_H^2}{4}\right)^{\frac{1}{2}}\sigma^*\right]$$

or

$$\dot{x}=\left[\left(\frac{\omega_p^2}{2}-\frac{\omega_H^2}{4}\right)^{\frac{1}{2}}+\frac{\omega_H}{2}\right]y=\left(\omega_0+\frac{\omega_H}{2}\right)y$$

$$\dot{y}=\left[\left(\frac{\omega_p^2}{2}-\frac{\omega_H^2}{4}\right)^{\frac{1}{2}}-\frac{\omega_H}{2}\right]x=\left(\omega_0-\frac{\omega_H}{2}\right)x,$$

where

$$\omega_0^2=\frac{\omega_p^2}{2}-\frac{\omega_H^2}{4}.$$

The orbits are given by $(\omega_0-\omega_H/2)x^2-(\omega_0+\omega_H/2)y^2=\text{constant}$. The following cases may be distinguished: for $\omega_0=0$ and $\omega_p^2=\omega_H^2/2$, the orbits are circles; for $0<\omega_0<\omega_H/2$ and $\omega_H^2/2<\omega_p^2<\omega_H^2$ (a familiar case), the orbits are ellipses with the ratio of axes given by $(\omega_H/2+\omega_0/\omega_H/2-\omega_0)^{\frac{1}{2}}$; for $\omega_0=\omega_H/2, \omega_p=\omega_H$, we have plane Brillouin or "slipping stream" flow; for $\omega_0>\omega_H/2, \omega_p>\omega_H$, the orbits are hyperbolic, the asymptotes having the slopes

$$\pm\tan^{-1}\left(\frac{\omega_0-\omega_H/2}{\omega_0+\omega_H/2}\right)^{\frac{1}{2}}.$$

It is clear from the form of the equation for σ that each flow may be regarded as a superposition of a circular flow and a hyperbolic one.

The equipotentials are always elliptical and are given by

$$V=-\frac{m}{2e}\left[\left(\omega_0-\frac{\omega_H}{2}\right)^2x^2+\left(\omega_0+\frac{\omega_H}{2}\right)^2y^2\right].$$

It would appear that by a suitable choice of applied voltages and electrodes one would be as likely to produce such flows as he is in the usual circular and linear arrangements.

It may be noted that if one attempts to extend the above results by looking for flows in which ρ is constant only along the individual stream lines, the only additional solutions appear to be those for the usual hollow circular Brillouin flow.

[1] For an account of Brillouin flow see J. R. Pierce, *Theory and Design of Electron Beams* (D. Van Nostrand Company, Inc., New York, 1949), p. 152.

Time Dispersion of Secondary Electron Emission*

E. W. Ernst† and H. VonFoerster
University of Illinois, Urbana, Illinois
(Received March 2, 1955)

IN this same space[1] the authors reported a method of producing and measuring short electron bunches approximately one micromicrosecond long. A simplified version of this technique was employed to determine the time dispersion of the secondary emission process by letting short electron bunches strike a secondary emission target and comparing the duration of the resulting secondary bunch with the measured duration of the primary bunch.

A schematic diagram of the experimental setup is given in Fig. 1. A fine electron beam, accelerated to about 3000 ev, is injected between the two wires of a Lecher wire pair,[2] excited at a frequency of 3000 megacycles, causing the beam to be deflected across a chopping slit which allows electrons to pass only during a small fraction of the rf period. These short segments of the beam—the primary bunches—enter a magnetic field region, labeled magnetic junction, where they suffer a deflection of exactly 90°, whereupon in a lens with short focal length (L_1) their speed is considerably reduced until they hit with perpendicular incidence the secondary emission target (platinum) with a velocity corresponding to about 500 ev. The secondaries produced by this impact travel in the opposite direction and will be accelerated by the same lens system (L_1) to enter the magnetic junction with an energy of approximately 2500 ev, whereupon they are deflected through a 90° arc in the other direction. Having left the junction, the secondary bunches are refocused (L_2) and an analysis is made of their duration. The analyzer[2]—the secondary uhf clock—consists of two uhf deflectors of the same type as used for the chopper, placed perpendicular to one another and excited such that during one period of the uhf signal, a dc beam would trace a complete circle on a fluorescent screen. Since these bunches will occupy only a fraction of one period of known duration, the observed segment on the screen in degrees is a measure of their duration (1°=0.925 $\mu\mu$sec). In order to determine the duration of the primary bunch, an identical uhf clock is placed closely behind the secondary emission target which easily slides in and out of the primary electron path. The experimental procedure is carried out in the three following steps: first, only the two uhf clocks are excited and adjusted until good approximations to circular traces on the two observation screens are alternately obtained with the secondary emission target in or out of position (Fig. 2); second, with the target removed and the circular deflectors in operation, power is increasingly fed into the chopper and its action observed on the primary screen until the two‡ chopped segments have been reduced sufficiently to give the desired duration of the primary bunches which are then

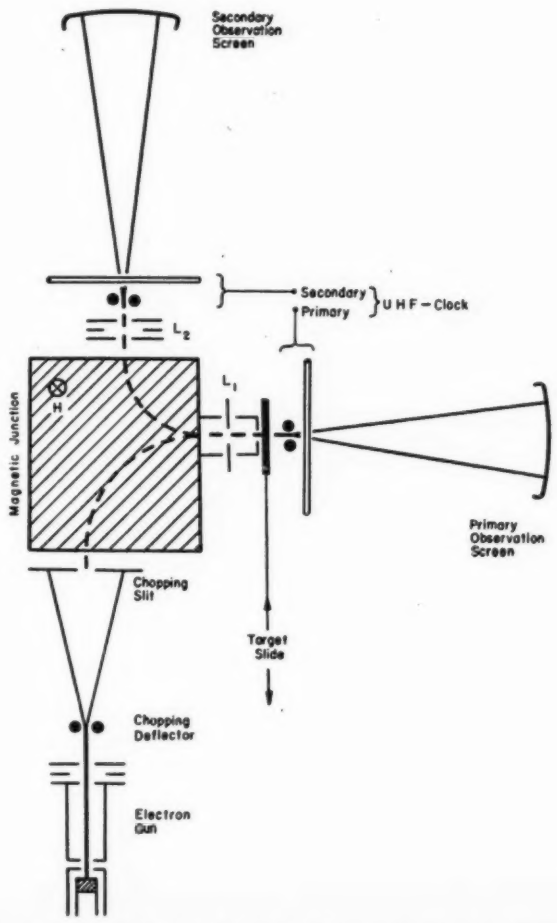


FIG. 1. Schematic diagram of a system for measuring the time dispersion of secondary emission.

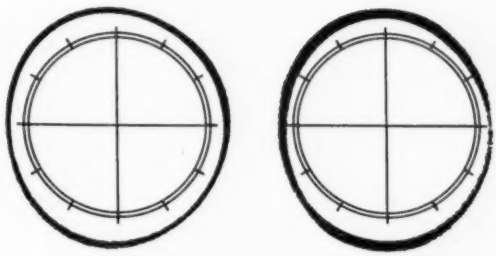


FIG. 2. Circular traces on the primary and secondary observation screen before chopping is applied. Deflection frequency 3000 Mc. One revolution, 333 $\mu\mu$sec.

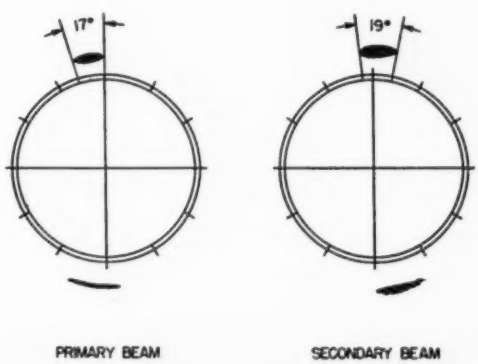


FIG. 3. Primary and secondary bunches while chopping is applied. By regulating the beam current I_0, the number of electrons in the primary bunches can be varied between 0.5 and 50 electrons per degree of the primary bunch (N/degree $=5.8I_0/\mu$A).

photographed; the third and final step is to slide the secondary emission target into position without changing the excitation of the chopping deflectors and to take a picture of the secondary observation screen.

A typical pair of traces is shown in Fig. 3 where the smaller segments on the primary and secondary screen correspond to a duration of 16 ± 2 $\mu\mu$sec and 18 ± 2 $\mu\mu$sec, respectively.

Since any system making use of the emission of secondary electrons will introduce aberrations of its own, it is clear that timing with this accuracy will indicate also the aberrations introduced by by the system itself. Consequently the total spread in bunch duration of 4 ± 2 $\mu\mu$sec, as evaluated from a large number of traces, will include the aberrations within the system, particularly those introduced by the initial velocity distribution of the secondary electrons. Hence, the figure of 6×10^{-12} second represents for a typical pure metal an upper limit of the time dispersion of the secondary emission process and is about one order of magnitude less than any results previously obtained by a direct method of measurement.[3,4]

Although this study has been directed toward the determination of the time element of a particular phenomenon, it is expected that the techniques will be applicable to other instances where time measurements of high resolution are necessary.

* This study is sponsored by the U. S. Air Force Cambridge Research Center.

† This material is based upon a dissertation submitted in partial fulfillment of the requirements for the Ph.D. Degree in Electrical Engineering at the University of Illinois.

[1] E. W. Grant and H. VonFoerster, J. Appl. Phys. **25**, 674–675 (1954).

[2] L. R. Bloom and H. VonFoerster, Rev. Sci. Instr. **25**, 649–653 (1954).

‡ The beam passes over the chopping slit twice in one period.

[3] M. H. Greenblatt, Phys. Rev. **95**, 632 (1954).

[4] *Note added in proof.*—A detailed account of M. H. Greenblatt's work has been published in the RCA Rev. **16**, 52 (1955).

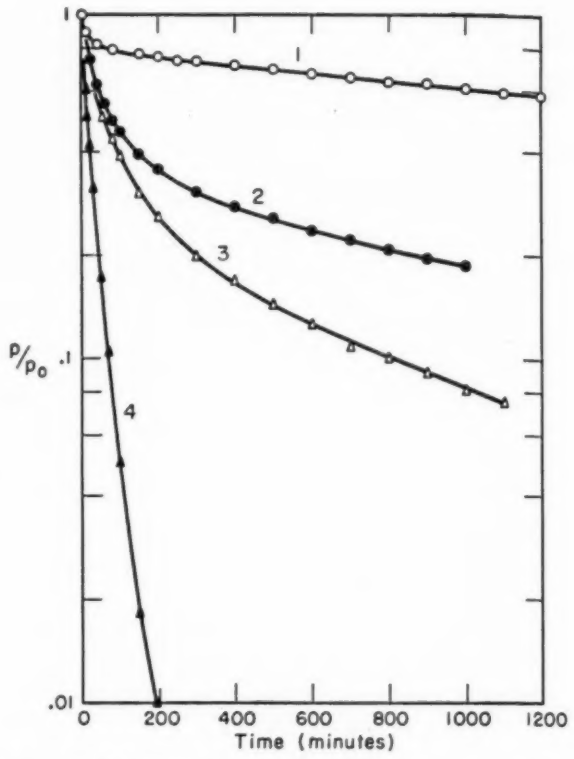


FIG. 1. Reduction of helium pressure in a vacuum system (normalized to p_0, the pressure at zero time) by pumping ionically with a Bayard-Alpert ionization gauge. Curves 1 through 4 represent runs after various degrees of outgassing of the gauge.

Ionic Pumping Mechanism of Helium in an Ionization Gauge

L. J. VARNERIN, JR., AND J. H. CARMICHAEL
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received March 2, 1955)

IONIC pumping or cleanup with a Bayard-Alpert ionization gauge[1] has been used to obtain ultrahigh vacua in suitably prepared vacuum systems. In these cases, helium, which diffuses through the glass from the atmosphere, is a major constituent of the gas pumped. While ionic cleanup is not well understood, the pumping of helium presents a particularly challenging case.

It has been believed generally that the gas is cleaned up at the glass walls of the tube. Bloomer and Haine[2] have presented some particularly convincing evidence on this score. In the case of helium, however, this explanation has been difficult to reconcile with other information. With the ion energies available (of the order of 150 ev), it is unlikely that many ions can be driven deeper than a few atomic layers. From the measured[3] values for the diffusion coefficient of helium in glass, simple diffusion calculations[4] show that a helium atom diffuses back through the surface in times of the order of microseconds and that only an insignificant amount of gas can be trapped.

We have suspected that the detailed surface condition of the glass may hold the explanation. In particular, in a vacuum processing the almost universal high temperature outgassing of the metal electrodes in tubes is likely to evaporate or sputter a slight amount of metal on the glass. Sometimes this film is evidenced by a slight light absorption. Often this film is not visible in carefully outgassed tubes. We propose that this film is responsible for ionic cleanup.

To test this hypothesis a Bayard-Alpert gauge which had no previous outgassing was used. A very mild heating of the grid by conduction current from one grid post to another evolved sufficient gas so that background gases were reduced to a tolerable level. The system containing the gauge (volume 1150 cm³) was filled with helium to an initial pressure of 8.5×10^{-5} mm Hg. The gauge was operated at 10 ma emission current (collector grounded, grid +180 volts, filament +30 volts). The reduction in pressure, normalized to the initial pressure p_0, is given as curve 1 in Fig. 1.

The grid and plate were then outgassed with increasing intensity to observe the effect on pumping. The outgassing was effected by electron bombardment; filament-grid potential being 750 volts ac, filament-plate potential, 375 volts ac. The amount of heating was controlled by the temperature of the filament.

Curve 2 of Fig. 1 was obtained after outgassing the tube structure for ten minutes at 700°C. No visible deposit on the glass was observed. However, an increased initial rate of pumping is observed followed by a slower long time pumping. A further increase in pumping is shown in curve 3 of Fig. 1 after a more intense outgassing for twenty minutes at 800°C. A very slight deposit may have been present. An estimated visible light absorption of 10 percent by the film leads to an estimate of film thickness of the order of 10 A. Curve 4 was obtained after a more intense outgassing at 1000°C. A clearly visible film was then present. This film was estimated to be at least 50 A thick. The significant increase in speed and capacity is to be noted. This type of curve is characteristic of tubes which are used to pump ionically to ultrahigh vacua.

As a further check that this metallic film is responsible for the observed pumping, a gauge with a film present was washed out with acid. Subsequent pumping curves resulted in a radical reduction in pumping.

It is now clear how the rapid pumping of helium can be explained. Ions driven into this metal film cannot diffuse as freely as in glass and can be trapped. For the very thin films the amount of gas which can be pumped depends upon the thickness of the film.

While the connection between the rapid pumping of helium and the metallic film on the glass walls of the gauge has been established, the small reduction in pressure in curve 1 is of some in-

terest. The use of a gauge in which the filament posts were covered by glass established that the major part of the gas is pumped into the filament posts. By maintaining the thin wire ion collector at grid potential between pressure readings, it was then established that a smaller part of the gas is pumped into the collector.

[1] D. Alpert, J. Appl. Phys. **24**, 860 (1953).
[2] R. N. Bloomer and M. E. Haine, Vacuum III **2**, 128 (1953).
[3] Rogers, Buritz, and Alpert, J. Appl. Phys. **25**, 868 (1954).
[4] T. Holstein, Westinghouse Research Memo 60-94411-9-20 (private communication).

Erratum: Analog Methods for Study of Transient Flow in Solids with Temperature-Dependent Thermal Properties

[J. Appl. Phys. **26**, 129–130 (1955)]

NORMAN E. FRIEDMANN
Department of Engineering, University of California, Los Angeles, California

THE symbolism of Fig. 1 is somewhat misleading. The actual terminal connection from R_2 is made to the output of an amplifier of gain $-A$ and not to C. The output of this amplifier is changed in sign, and then the outside feedback loop is closed through C.

Books Reviewed

The Elements of Probability Theory. HARALD CRAMÉR. Pp. 281. John Wiley and Sons, Inc., New York, 1955. Price $7.00.

This book provides an introduction to probability theory and its applications. It is intended as a text for upperclass undergraduate students and as a reference book for engineers, biologists, and others who wish to apply statistical methods. The book begins with the fundamental definitions and rules for calculating probabilities. Part II presents probability distributions and random variable theory. Applications are considered in Part III; some of the topics in this part are sampling distributions, statistical inference, and the χ^2 test.

Abstracts of the Literature on Semiconducting and Luminescent Materials and Their Applications. Compiled by the staff of the Battelle Memorial Institute. Pp. 169+x, 1953 issue. John Wiley and Sons, Inc., New York, 1955. Price $5.00.

This volume of abstracts of the 1953 literature was written by the Solid State Devices Division of Battelle under the sponsorship of the Electrochemical Society. There are 775 articles listed, and abstracts are provided for almost all of these. The volume is organized by materials (e.g., germanium, silicon, sulfides) except for sections on luminescence and semiconductor theory.

Strength and Resistance of Metals. JOHN M. LESSELS. Pp. 450+xiv, John Wiley and Sons, Inc., New York, 1954. Price $10.00.

The aim of this book is to provide information on the behavior of metals under stress. The book is intended for seniors, graduate students, and design engineers. Although most of the material presented is concerned with steel, nonferrous alloys are also discussed. Tension testing and tension properties, at ordinary and elevated temperatures, are described first. Hardness, impact, fatigue, fracture, strain hysteresis, mechanical wear, and theories of strength are the subjects of subsequent chapters. There are about 200 problems for the student.

Storage Batteries. GEORGE WOOD VINAL. Pp. 446+xi. Fourth Edition. John Wiley and Sons, Inc., New York, 1955. Price $10.00.

The new fourth edition of this book represents a considerable revision of the 1940 edition. About half the figures are new and the text has been largely rewritten. The physical and chemical properties of the materials of lead-acid, silver oxide, nickel-iron, and nickel cadmium storage cells are described. The electrical properties and testing of batteries are discussed. The final chapters present applications of storage batteries.

What Every Engineer Should Know About Rubber. W. J. S. NAUNTON. Pp. 128, British Rubber Development Board, London, 1954. Order from Natural Rubber Bureau, 1631 K Street, N.W., Washington 6, D. C. Price $0.50.

The aim of this small, clothbound book is to present the properties and applications of natural rubber. The book is intended primarily for the design engineer and is illustrated with 145 photographs and line drawings.

Applied Mass Spectrometry. Pp. 333+vii. Institute of Petroleum, London, 1954. Price £2–10–0.

A conference on applied mass spectrometry was held in London in October, 1953. It was sponsored by the Institute of Petroleum. This book presents the papers given at that conference. There are 26 papers in the following areas: analytical applications, applications to fundamental problems, instrument development, and computing methods. A bibliography covering the period September, 1950 to June, 1953, lists 400 references.

Tables of Functions and of Zeros of Functions. National Bureau of Standards Applied Mathematics Series No. 37. Pp. 211+ix. Order from Government Printing Office, Washington 25, D. C. Price $2.25 (foreign: $3.00).

This clothbound volume is the first of a projected series which will bear the subtitle "Collected Short Tables of the NBS Computation Laboratory." The present volume collects 15 short tables originally published in 5 different mathematics journals and also presents 3 tables which have not previously been published. There are 10 tables of special functions, such as integrals of the Bessel functions J_0 and Y_0, exponential integrals, Struve functions, values of $x^n/n!$, and 8 tables of zeros of functions. The latter group includes zeros of the Legendre and Laguerre polynomials together with the weight factors needed for quadrature, as well as zeros of various kinds of Bessel functions. An introduction preceding each table gives a discussion of the theory underlying the table and details of method of computation, bibliographic background, and numerical illustrations. An introduction to the whole volume describes the scope of the various tables and incorporates, as a special feature, a reference to the principal review of each table.

Tables of the Error Function and Its Derivative. National Bureau of Standards Applied Mathematics Series No. 41. Pp. 302+xi. Order from Government Printing Office, Washington 25, D. C. Price $3.25 (foreign: $4.35).

The functions

$$(2/\pi^{\frac{1}{2}})\int_0^x \exp(-t^2)dt \quad \text{and} \quad (2/\pi^{\frac{1}{2}})\exp(-t^2)$$

are tabulated with increments of 0.0001 in x for $0<x<1$ and with increments of 0.001 for $1<x<5.6$. This clothbound volume is a re-issue of "Mathematical Table 8" which was first published in 1941.

Tables of Sines and Cosines for Radian Arguments. National Bureau of Standards Applied Mathematics Series No. 43. Pp. 278+xi. Order from Government Printing Office, Washington 25, D. C. Price $3.00 (foreign: $4.00).

The functions $\sin x$ and $\cos x$ are tabulated to 8 places for values of x in radians between 0 and 25 and increments of 0.001. There are a few pages of supplementary tables of $\sin x$ and $\cos x$ for very large and very small arguments.

Electricity and Magnetism. RALPH P. WINCH. Pp. 755+xi. Prentice-Hall, Inc., New York, 1955. Price $7.75.

This text is intended to be used in a one-year course in electricity and magnetism at the sophomore or junior level in a college or university. The calculus is used. Vector algebra and the gradient vector are explained where they are first used so the student need have no previous knowledge of these subjects. First considered are dc and ac circuits and circuit elements. Electrostatic and magnetostatic fields are the subject of the middle one-third of the book. Induced emf's, magnetic properties of materials, and oscillators and radiation are discussed in the last one-third. Rationalized MKS units are employed throughout the book, but important quantities are also given in cgs units. There are about 400 problems for the student.

Advances in Electronics and Electron Physics. Vol. VI. EDITED BY L. MARTON. Pp. 538+xi. Academic Press, Inc., New York, 1954. Price $11.80.

This sixth volume in the series has an addition (*and Electron Physics*) to the title in order to describe more adequately the subjects covered by the series. The physics of electrons in solids is given as much attention in the present volume as the physics of electrons in vacuum or gases. The individual critical reviews presented are: "Metallic Conduction at High Frequencies and Low Temperatures," by A. B. Pippard; "Relaxation Processes in Ferromagnetism," by E. Abrahams; "Ferrites," by J. Van den Handel; "Comparison of Semiconductor and Gaseous Electronics Devices," by W. M. Webster; "Electron Microscope," by M. E. Haine; "Traveling-Wave Tubes," by R. G. E. Hutter; and "Space Charge Limited Currents," by H. F. Ivey.

Books Received

Table of Salvo Kill Probabilities for Square Targets. National Bureau of Standards Applied Mathematics Series—44. Issued December 20, 1954. Pp. 33. Price $0.30.

Experiments in the Computation of Conformal Maps. National Bureau of Standards Applied Mathematics Series—42. Issued January 20, 1955. Pp. 61.

Applied Mass Spectrometry. Institute of Petroleum, London, W. 1, 1954. Pp. 333, with Figs.

The Elements of Chromatography. TREVOR I. WILLIAMS. Pp. 90, Figs. 36. Philosophical Library, New York, 1955. Price $3.75.

Advances in Electronics and Electron Physics. Vol. VI. L. MARTON. Pp. 538+xi, with Figs. Academic Press, Inc., New York, 1954. Price $11.80.

Bibliography on Hearing. DOROTHY COHEN, Technical Editor. Pp. 599. Harvard University Press, Cambridge, Massachusetts, 1955. Price $7.00.

Wheeler Monographs, Vol. I. HAROLD A. WHEELER. Wheeler Laboratories, Great Neck, New York, 1953.

Spectros Molecular. II. (in *Interlingua*). S. M. FERIGLE AND ALFONS WEBER. Pp. 27+v. Spectroscopy Laboratory, Illinois Institute of Technology, Chicago 16, Illinois, c/o Forrest F. Cleveland, Editor, 1955. Price $1.00.

A List of American Standards. Pp. 48. 1955 edition. Available free from the American Standards Association, 70 East 45th Street, New York 17, New York.

Public Relations in Education. BROWNELL-GANS-MAROON. Pp. 249+xi. McGraw-Hill Book Company, Inc., New York, 1955. Price $4.50.

Science and the Human Imagination. MARY B. HESSE. Pp. 171. Philosophical Library, New York, 1955. Price $3.75.

Servomechanisms and Regulating System Design. Vol. II. HAROLD CHESTNUT AND ROBERT W. MAYER. Pp. 384+xiii, with Figs. John Wiley and Sons, Inc., New York, 1955. Price $8.50.

Research Reports

Axial Variation of the Magnetic Field in Solenoids of Finite Thickness. D. E. Mapother and J. N. Snyder. Circular No. 66, 70 pages. Write to University of Illinois Engineering Experiment Station, 112 Civil Engineering Hall, Urbana, Illinois. Price: Free until September 1, 1955; $1.00 thereafter.

This work has been prepared to serve as a practical handbook for the design of solenoids where the effect on the magnetic field of the finite thickness of the solenoid windings must be considered. A semiempirical approach is adopted which puts considerably less strain on the designer's mathematical proficiency than the analytical methods which have been described in the literature. The work on which this report is based was done in the University of Illinois Physics Department.

A simple correction function is defined which makes it possible to express the axial field of a thick solenoid in terms of the elementary formula for the axial field of an ideal solenoid (having infinitesimal winding thickness). By use of the facilities of the University of Illinois High Speed Electronic Digital Computer (the ILLIAC), this correction function has been tabulated for an extensive range of solenoid lengths and cross-sectional dimensions. Corresponding tabulations for loops of various relative thickness are also included, a loop being defined in the present instance as a solenoid having a square winding cross section. Within the range of tabulated values (which is believed to cover a large majority of practical situations), the designer can obtain very detailed information about the magnetic field with practically no calculation at all. The tables are also useful in giving a quantitative impression of the importance of the thickness correction for increasing thickness of winding as well as for position along the axis.

Thermionic Emission from Semiconductors. E. O. Kane. December 31, 1954, 95 pages. Write to Department of Physics, Cornell University, Ithaca, New York. Price: $1.00.

The emphasis in this report is on the theory of thermionic emission from semiconductors. Some introductory experiments on emission from single crystals of barium oxide are described and compared with theory, but most of the report is concerned with the theory of emission from any n-type semiconductor. The effects on emission of an applied electric field and of the flow of current are discussed, and the limitations of the equilibrium treatment are considered.